Stat5 Binding to Chromatin

Gillian H. Little

PhD Thesis

University of Edinburgh

2004

Declaration

I declare that the work presented in this thesis is my own and that the contribution of others has been clearly indicated. This work has not been submitted for any other degree or qualification.

Gillian H. Little

Abstract

Nucleosomes often sit at precisely defined positions on eukaryotic gene promoters, influencing the regulation of target genes. Expression of milk proteins including β-lactoglobulin is controlled by prolactin activation of the transcription factor Stat5 via the Janus kinase / Signal transducer and activator of transcription (Jak/Stat) pathway. Stat5 has previously been shown to tetramerise where binding sites are tandemly linked and the proximity of these binding sites appears to be important for these interactions. This work and previous large scale mapping of the β-lactoglobulin promoter shows that the dyad of a strongly positioned nucleosome lies at –184 bp from the transcription start on the promoter of the β-lactoglobulin gene. This brings together two binding sites for Stat5, at the points of entry and exit of DNA from the nucleosome that would otherwise be spaced 185 bp apart, an arrangement that could potentially bring bound Stat5 dimers close enough to facilitate tetramerisation. The chromatin structure over the active and inactive gene promoter is different; there are two alternative nucleosome positions in the active and only one in the inactive promoter. One of these positioning sites would not allow the tetramerisation interaction to take place. In order to understand better the mechanisms by which the expression of β-lactoglobulin is regulated by Stat5 we set out to investigate the role of these positioned nucleosomes in Stat5 binding *in vitro*. Stat5A and B binding patterns on both naked DNA and on reconstituted chromatin probes are shown by a series of bandshift experiments using purified recombinant Stat5 produced in a baculovirus expression system. Characterisation of Stat5 reveals the protein to be phosphorylated and able to bind DNA. A mutation, W37A, which removes the ability of Stat5 to form dimer-dimer interactions was employed to further investigate a potential role of tetramerisation influencing Stat5 binding in a chromatin context. This architectural feature could act to control the temporal and tissue specific expression of β–lactoglobulin.

Acknowledgements

I would like to start by thanking my two supervisors, Bruce and Jim, for their enthusiasm and insight into my project. Thanks must also go to the members of the Allan and Whitelaw, labs both past and present, for pointing me in the right direction and to everyone in the South wing who has come to my aid and who has been involved in the general running of the lab.

I would not have got far without the kind donations of donor chromatin from Jim Allan and the Stat5 DNAs from W. J. Leonard. Eve Devinoy, Benjamin Millot and Christine Watson provided help with getting the Stat5 bandshifts up and running, and the production of recombinant Stat5 was aided by the use of spinner culture equipment from Gordon Allan. A COST short term scientific mission grant made possible my visit to the Devinoy lab.

And of course thanks must go to all the inhabitants, past and present, of the Roslin student room, to the Whitelaw and Allan lab members and to all my friends and family for looking after my social, physical and mental well-being.

Y'all know who you are!

Contents

List of Figures

List of Tables

List of Abbreviations

AcMNPV	*Autographa californica* Multiple Nuclear Polyhedrosis Virus
APS	Ammonium persulphate
BEVS	Baculovirus Expression Vector System
BLG	β-lactoglobulin
BME	β-mercaptoethanol
bp	Base pairs
BSA	Bovine serum albumin
ChIP	Chromatin immunoprecipitation
CIAP	Calf intestinal alkaline phospatase
cpm	Counts per minute
cps	Counts per second
dH_2O	Deionised water
D/I/P	Dexamethasone, insulin and prolactin
dNTPs	Deoxyribonucleoside triphosphates
DTT	Dithiothreitol
EDTA	Ethylene diaminetetra-acetic acid
EGF	Epidermal growth factor
EMSA	Electrophoretic mobility shift assay
EPO	Erythropoietin
ER	Estrogen receptor
ERE	Estrogen responsive element
ExoIII	Exonuclease III
ffu	Focus forming units
GAS	Gamma activating sequence
GH	Growth hormone
GHR	Growth hormone receptor
GM-CSF	Granulocyte-macrophage colony stimulating factor
GR	Glucocorticoid receptor
HAT	Histone acetyl transferase
HP1	Heterochromatin protein 1
HRP	Horseradish peroxidase
HS	DNaseI Hypersensitive site
IFN	Interferon
IL	Interleukin
Jak	Janus Kinase
kb	Kilo base pair
kda	Kilo Daltons
LM PCR	Ligation mediated PCR
MAR	Nuclear matrix attachment region

MCS	Multiple Cloning Site
MGF	Mammary gland factor
MNase	Micrococcal Nuclease
MMTV LTR	Mouse mammary tumour virus long terminal repeat
MOI	Multiple of infection
MPBF	Milk protein binding factor
nt	Nucleotides
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEV	Position effect variegation
pfu	Plaque forming units
pp	Postpartum
rStat5	Recombinant Stat5
SF9	Spodoptera frugitera derived cell line
SH2	Src homology 2
SOCS	Suppressors of cytokine signalling
Stat	Signal transducers and activators of transcription
T4 PNK	T4 polynucleotide kinase
TAE	Tris, acetic acid, and EDTA.
TBE	Tris, boric acid and EDTA
TBS	Tris buffered saline
TEMED	N,N,N',N'-Tetramethyl-ethylenediamine
TPO	Thrombopoietin
U	Units
WAP	Whey acidic protein
W37A	Tryptophan 37 to alanine mutation
WT	Wild Type
YY1	Ying Yang 1

1 INTRODUCTION

Differentiated cells throughout the body are controlled by the precisely regulated expression of genes. During development and all cell functions genes are switched on and off at precisely regulated time points, so that although every cell contains the same DNA sequences, only the correct genes are active at any time. Whether a particular gene will be on or not at a given time in a specific tissue is controlled by a complex interaction of transcription factors with DNA and its associated structural elements. Transcription factors can act in a variety of ways: by modifying the DNA, by altering chromatin structure or by recruiting or removing other factors including the RNA polymerase complex. They may be regulated by external signals, or by gene products or chromatin structures already present inside the cell.

But how are all these genes regulated in so specific a manner? Each gene has regulatory regions associated with it that can bind different combinations of transcription factors. Some of these may be common to many genes, such as general transcription factor sites like the TATA box. Others are more specialised, involved in the regulation of perhaps only a set of related genes.

1.1 Chromatin.

Most studies of transcription factors have historically involved binding to naked DNA, but in a eukaryotic cell this is not the native state of DNA. Here DNA is organised into a series of higher order structures called chromatin. The role of chromatin is to maintain DNA in a highly ordered compact structure, and it is to this substrate that transcription factors must bind in order to affect expression of a gene. Only recently have techniques such as

chromatin immunoprecipitation (ChIP) and nucleosome mapping at high resolution been developed for the study of transcription factors *in vivo*.

The evolution of the nucleus, and the organisation of DNA into chromatin, marks a major difference between prokaryotes and eukaryotes, as does the presence of repetitive elements in eukaryotic genomes requiring a higher degree of organisation of the DNA (Hickey, 1992; Bowen and Jordan, 2002). Some theories state that it is the organisation of DNA into nucleosomes or nucleosome-like structures that marks the real phylogenetic branch between prokaryotic and eukaryotic organisms (Bendich and Drlica, 2000) but this is by no means certain. Archaeabacteria, which are classified somewhere between eukaryotes and prokaryotes contain structures very similar to the eukaryotic nucleosome, but have not evolved the nuclear membrane (Woese and Fox, 1977; Shioda et al., 1989; Pereira et al., 1997). Whatever the barrier the increased level of regulation available, brought in when transcription and translation were physically separated and DNA was packaged in this ordered manner, allowed for the development of much larger genomes which then could give rise to more complex organisms. Most importantly cells could differentiate and take on a specific role within a multicellular organism. This required the development of mechanisms to overcome the inherent repressive nature of the organisation of DNA into chromatin. Mechanisms also evolved to take advantage of the nucleosomal structure of DNA in the specific regulation of genes (Struhl, 1999).

1.1.1 Euchromatin and Heterochromatin.

Chromatin can be divided into two basic states, which were first described based on the uptake of dye in different areas of the nucleus (Heitz, 1928). These are heterochromatin, a highly condensed form of DNA, and

euchromatin, which is more loosely packed. In general, a gene situated in a region of heterochromatin will be repressed in comparison to genes in euchromatic regions (Grewal and Elgin, 2002; Vermaak et al., 2003). This is demonstrated in the phenomenon of position effect variegation (PEV), where an identical transgene may show different expression levels depending on the site of integration (Dobie et al., 1997). This can be partly explained for by the enhanced access of transcription factors to the more open chromatin structure if the transgene integrates into a euchromatic region rather than a region of heterochromatin. Inclusion of barrier elements in transgenes can overcome PEV (Sun and Elgin, 1999; Festenstein and Kioussis, 2000). Such barriers stop the spread of heterochromatin structure into a transgene.

Heterochromatin is characterized by a closed chromatin structure that is resistant to cleavage by nucleases, by the presence of heterochromatin protein 1 (HP1), and by characteristic histone modifications such as methylation of histone H3 lysine 9 (H3-K9) and hypoacetylation of histone tails (Struhl, 1998; Lusser, 2002; Grewal and Elgin, 2002). Recently the discovery was made that RNA plays an important role in the maintenance of the heterochromatin state (Volpe et al., 2002).

1.1.2 Nucleosomes.

The fundamental repeating unit of chromatin, the core particle, is 146 base pairs of DNA wrapped in 1.7 helical turns wound round a core histone octamer containing one copy each of an H3 and an H4 dimer and 2 copies of an H2A:H2B dimer. The structure of the nucleosome core particle is known in detail (Luger et al., 1997; Richmond and Davey, 2003). A linker histone such as H1 joins the complex to form a functional chromatosome, increasing the DNA protected against nucleases to 160bp (Simpson, 1978). A schematic

diagram of nucleosomes and a representation of how higher order chromatin structures might form is shown in Figure 1.1.

The next stage in chromatin organisation is the 30nm fibre, formed from higher order interactions within the 10nm fibre in a process that is probably stablilised by linker histone interactions (Carruthers et al., 1998). The 30nm fibre then folds to form further higher order structures, culminating in the formation of chromosomes. The DNA is eventually compacted by a factor of 50 000, fitting over a metre of DNA into the nucleus of a cell less than 10μm across. An overview of the basic forms of chromatin structure is shown in Figure 1.1. Little is known about the structures formed in the 30nm fibre and further higher order structures.

1.1.3 Nucleosome Modifications.

The amino-terminal tails of the core histones can be covalently modified. Lysine residues can be mono-, di- or tri-methylated in what appears to be a relatively static mark, and acetylated or phosphorylated in a dynamic process that may encode transcriptional information. Arginine residues can be methylated and serine residues phosphorylated. These modifications may be recognised by regulatory factors as a histone code signal as part of transcription regulation (Jenuwein and Allis, 2001; Lusser, 2002). For example the chromodomain of (HP1) specifically recognises the methylated form of H3-K9 in the repressed chicken β-globin locus resulting in a heterochromatin structure. In the active locus H3-K4 is methylated and H3 and H4 are acetylated (Kim and Dean, 2004). Histone acetyl transferases (HATs) such as CREB binding protein (CBP)/p300 acetylate histone tails. Conversely, histone deacetylases (HDACs) remove these acetyl groups

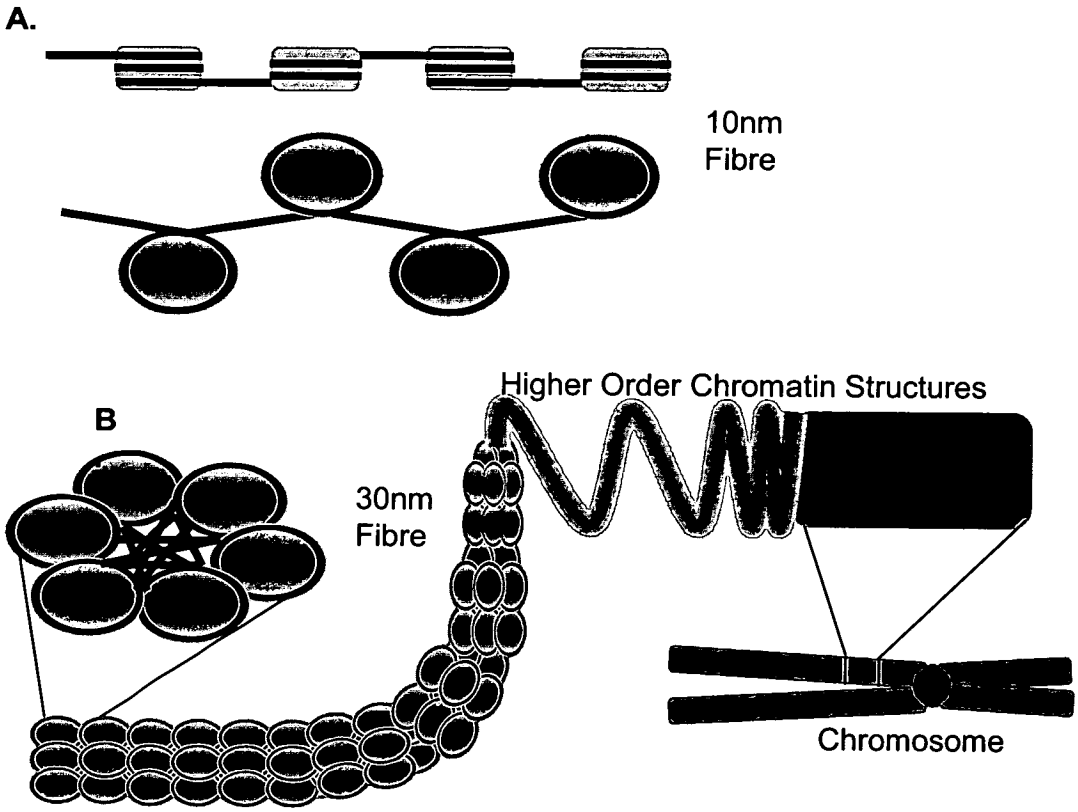


Figure 1.1. Chromatin Structure from Nucleosomes to Chromosomes.

Schematic diagram showing how **A**. DNA (black) is organised into a left handed helice round nucleosomes (grey) in the 10nm fibre, and **B**. how this may fold up to give the 30nm fibre and other higher order chromatin structures.

resulting in a more closed chromatin structure (Struhl, 1998; Wolffe, 2001; Lusser, 2002). As well as acting as a mark on chromatin, the addition of acetyl groups neutralises the positive charge of the lysine residues resulting in less tight interactions of the histone tails with the DNA and therefore a more open chromatin structure. The presence of these mechanisms serves to emphasise the important role nucleosomes have in gene regulation.

1.1.4 Nucleosome Movement.

ATP-dependent nucleosome remodelling complexes such as SWI/SNF use energy from the hydrolysis of ATP to move nucleosomes on DNA (Becker and Horz, 2002). These and other non-specific factors are recruited by binding to sequence specific factors, and can be involved in both activation and repression of transcription. It is likely that part of the activity of these complexes is to remove one or both of the H2A/H2B dimers (Workman and Kingston, 1998; Bruno et al., 2003). Nucleosomes also exhibit a tendency to redistribute on a DNA fragment in a temperature dependent manner that does not require ATP (Pennings et al., 1991; Meersseman et al., 1992).

1.2 Nucleosome Positioning.

Nucleosomes have been shown to sit at precisely regulated positions on DNA (Buckle et al., 1991; Simpson, 1991; Thoma, 1992). They are probably located there by sequence signals in the DNA, but other factors such as proteins bound to DNA may have additional influences. Nucleosome positions have been mapped to base pair accuracy *in vitro* in several eukaryotic genes, including the chicken β-globin gene (Buckle et al., 1991; Kefalas et al., 1988) and the ovine β-lactoglobulin (BLG) gene and promoter (Boa, 1999; Gencheva and Allan, 2005), using various techniques including the monomer extension technique, Exonuclease III and restriction mapping

of core particle DNA (Yenidunya et al., 1994; Meersseman et al., 1991; Kefalas et al., 1988).

1.2.1 Nucleosomes and Gene Regulation

Generally, the presence of nucleosomes on DNA is regarded as having an inhibitory effect on gene expression, as interactions of DNA with histones restricts access for transcription factors, many of which cannot bind to sites on a nucleosome. Although nucleosomes are dynamic structures and DNA sequences throughout much of the core can become transiently dissociated from histones, only a limited number of base pairs are available at any one time. The curve of DNA organised in a nucleosome is such that many factors do not recognise their binding sites, the rotational setting of DNA on a nucleosome can also affect whether a factor will bind (Martinez-Campa et al., 2004). As such, the presence of nucleosomes on regulatory regions can have a fundamental role in regulation of transcription, although the precise effect varies between genes. Nucleosomes on regulatory regions may be either displaced or remodelled on gene activation. Once formed the RNA polymerase II complex can transcribe through bound nucleosomes (Felsenfeld et al., 1996).

Nucleosomes do not necessarily have a negative impact on transcription. There are examples of the opposite occurring where the winding of DNA round a positioned nucleosome brings sequences together to allow interactions between two bound transcription factors, effectively removing the sequence between the sites (Jackson and Benyajati, 1993; Schild et al., 1993; Zhu and Thiele, 1996; Stünkel et al., 1997). Another model is based on the formation of a "supergroove" where sites separated by 80bp and

organised in a nucleosome are brought together (Edayathumangalam et al., 2004).

1.2.2 Interaction of Transcription Factors with Nucleosomal DNA.

Transcription factors may or may not be able to bind within chromatin. Partly this relies on the location of binding sites either within linker DNA or within DNA associated with a nucleosome (Figure 1.2). Certain transcription factors do not appear to bind at all to a chromatin template, unless it has been remodelled (Spangenberg et al., 1998). Other transcription factors can bind to DNA associated with a nucleosome, e.g. HNF3 binds to DNA at the periphery of a nucleosome on the vitellogenin promoter (Robyr et al., 2000). The DNA sequences at nucleosome boundaries bind the histone octamer less tightly and are therefore more accessible than more central sequences (Weischet et al., 1978). Restriction enzyme access to nucleosomes is 1000 times less in the central 100bp and 50 times less in end sequences compared to naked DNA (Linxweiler and Horz, 1984; Anderson et al., 2002).

The MMTV promoter has been extensively studied with regard to nucleosome positions and transcription factor access therein, notably of the glucocorticoid receptor (GR) and nuclear factor 1 (NF-1) (Richard-Foy and Hager, 1987; Perlmann and Wrange, 1988; Cordingley et al., 1987; Piña et al., 1990; Eisfeld et al., 1997; Ostland Farrants et al., 1997; Spangenberg et al., 1998). GR will bind to DNA associated with the periphery of a positioned nucleosome, but not to sites closer to the dyad. GR binding results in nucleosome remodelling which allows NF-1 to bind; NF-1 will not bind to DNA associated with a nucleosome, but it will bind to a nucleosome lacking H2A/H2B. Modification of nucleosomes in this way has been suggested to be one of the actions of chromatin remodelling complexes such as SWI/SNF

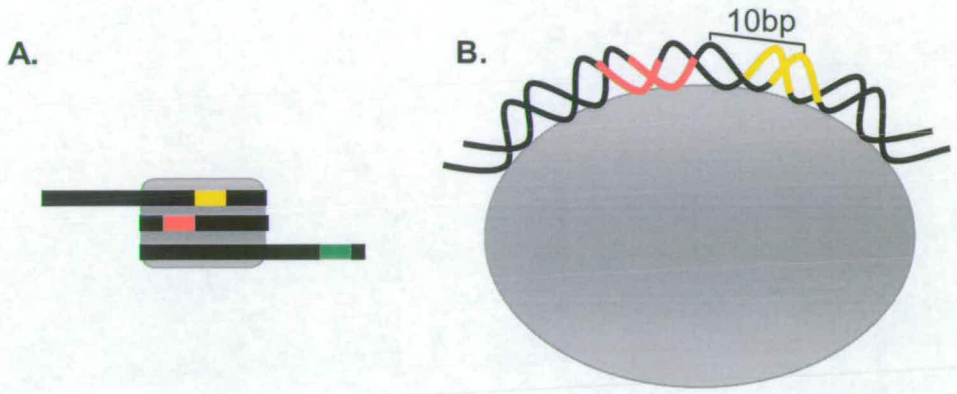


Figure 1.2 Nucleosome Structure: Transcription Factor Access.

A. Binding of transcription factors to sites on DNA can be affected by their interactions with nucleosomes. Sites in the linker DNA (green) are generally accessible for binding but sites in a nucleosome may or may not be accessible. Sites close to the nucleosome boundary (amber) are more likely to be accessible than those more centrally placed (red). **B.** The rotational setting of the DNA can also affect transcription factor binding. Factors may be able to access binding sites facing away from the nucleosome core (amber) but not those facing towards it (red), for example this is how the 10bp periodicity of DNaseI digestion on a positioned nucleosome comes about.

(Section 1.1.4). At another promoter, the rat tyrosine aminotransferase, GR binding results in removal of the nucleosome rather than remodelling (Flavin et al., 2004) indicating that the behaviour of nucleosomes is by no means general.

The organisation of DNA into nucleosomes can have a positive effect on transcription. The *Xenopus* vitellogenin B1 promoter, the *Candida. glabrata* metal responsive promoter, the *Drosophila* alcohol dehydrogenase (adh) promoter and the human U6 gene are examples of the wrapping of DNA round a nucleosome bringing together regulatory elements resulting in enhanced gene expression (Jackson and Benyajati, 1993; Schild et al., 1993; Zhu and Thiele, 1996; Stünkel et al., 1997). The mechanism of this action in the adh promoter and in the vitellogenin promoter are illustrated in Figure 1.3.

1.2.3 Positioning Signals.

Precisely what makes nucleosomes sit at defined positions on DNA is still a matter for debate. Reconstitutions onto naked DNA *in vitro* can result in precisely positioned nucleosomes, which may sit at similar positions to *in vivo* mapped nucleosome positions (Jackson and Benyajati, 1993; Boa, 1999; Hansen and van Holde, 1991; Linxweiler and Horz, 1985). The $(H3H4)_2$ tetramer is responsible for positioning nucleosomes (Prunell, 1983). The inherent bendability of the DNA sequence seems to be an important factor (Jackson and Benyajati, 1993) as the DNA wound round the nucleosome is heavily bent/twisted and contains several kinks (Hogan et al., 1987). CpG methylation of DNA may provide an additional factor to consider in directing nucleosome positioning (Davey et al., 1995; Davey et al., 2003).

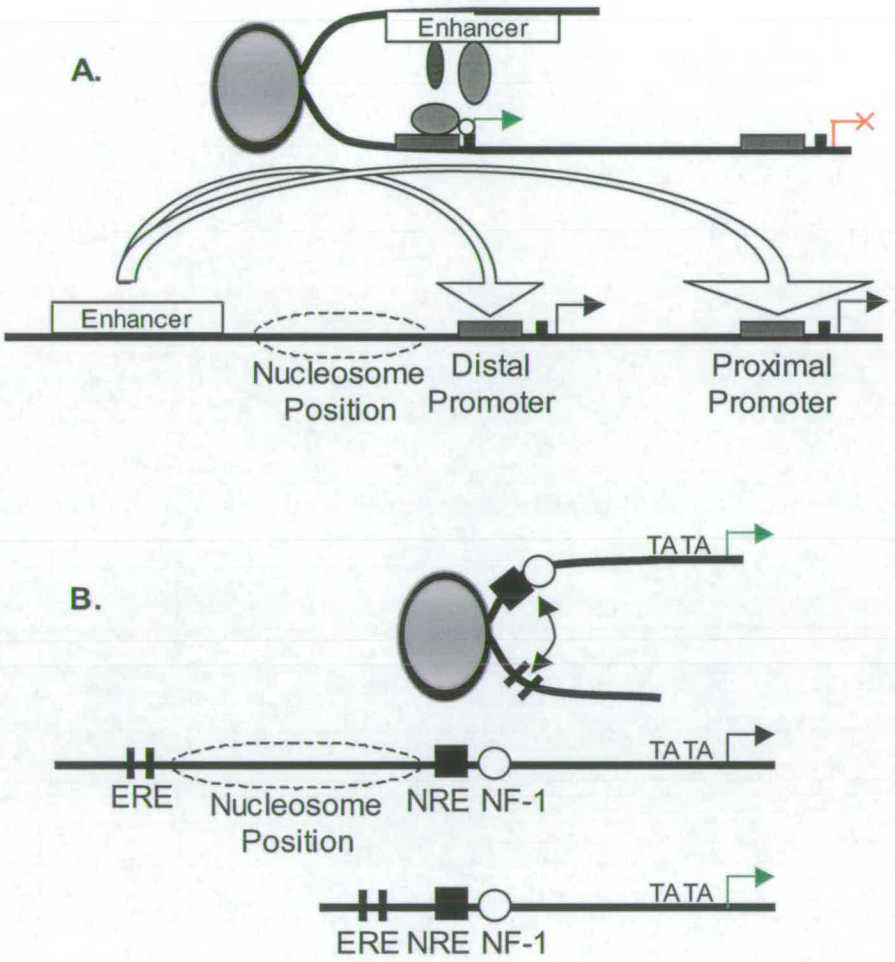


Figure 1.3 Models of Positioned Nucleosomes Enhancing Transcription Factor Interactions.

A. A schematic representation of the *Drosophila* alcohol dehydrogenase promoter, adapted from (Jackson and Benyajati, 1993). In the absence of a positioned nucleosome, the enhancer can interact with either the proximal or the distal promoter. A tissue-specific positioned nucleosome ensures the enhancer is much more likely to interact with the distal promoter in fat body tissue. **B**. A schematic representation of the *Xenopus* vitellogenin B1 promoter. The positioned nucleosome brings together an estrogen responsive element (ERE), with the proximal promoter thus facilitating enhanced transcription (top). Deleting the nucleosome sequence (bottom) has the same effect. This diagram is adapted from (Schild et al., 1993).

1.2.4 Nucleosome Reconstitution.

Nucleosomes can be reconstituted onto naked DNA sequences *in vitro* by the transfer of histones from donor chromatin or from purified core histones by salt gradient dialysis to give a structure indistinguishable from native nucleosomes (Tatchell and van Holde, 1977; Riley and Weintraub, 1978; Luger et al., 1997). At high salt concentrations the histone core is dissociated from DNA, but exists as an octamer (Thomas and Kornberg, 1975). As the salt is removed histones bind to DNA. In the absence of DNA the octamer will also become unstable and dissociate into $(H3H4)_2$ and H2A-H2B tetramers (Tatchell and van Holde, 1977). Statistically a stretch of DNA more favourable to being organised into a nucleosome will have nucleosomes associated with it more often. This is referred to as the strength of a nucleosome position. As [NaCl] is brought down to ~1M, a histone tetramer $(H3H4)_2$ forms and binds (Wilhelm et al., 1978). The position of the tetramer defines the eventual location of the nucleosome. The core histone octamer is partially formed at 800mM NaCl, and at 600mM NaCl assembly of nucleosomes is complete, although not conformationally stable until in a low ionic strength buffer (Tatchell and van Holde, 1977; Hansen and van Holde, 1991; Dong and van Holde, 1991).

1.3 Mapping Nucleosome Positions.

Many positioned nucleosomes exhibit a group of rotationally positioned nucleosomes, spaced 10bp apart (Dong et al., 1990; Meersseman et al., 1991; Pennings et al., 1991). The rotational position of a nucleosome can be determined by accessibility to nucleases or chemical cleavage reagents to the side of the DNA helix that faces into or away from the nucleosome (Figure 1.2b). DNaseI, which attacks DNA in the minor groove, produces a characteristic 10bp pattern in digests of a rotationally positioned nucleosome

as the minor groove is alternatively facing towards and away from the histone core. The presence of a group of nucleosomes with a spacing other than 10bp will not show up in a DNaseI digest, as the cutting patterns cancel each other out, resulting in a digest pattern similar to that of naked DNA.

The translational position of a nucleosome is defined as the position of a nucleosome on the DNA from a set point and can be mapped by the protection of DNA covered by the nucleosome from attack by nucleases. The monomer extension technique (Yenidunya et al., 1994) makes use of core DNA protected from MNase digestion by nucleosomes reconstituted onto the sequence of interest. Isolated core DNA is annealed back to a single stranded version of the template, and extended to a known restriction site. Extension products are resolved on a sequencing gel and from their size nucleosome boundaries can be mapped with base pair accuracy. Another *in vitro* nucleosome mapping technique makes use of the stepwise digestion of the 3′ strand by ExoIII. ExoIII digestion will pause at a nucleosome boundary, although eventually it will also digest internally to the nucleosome pausing at 10bp intervals as DNA / histone interactions occur (Riley and Weintraub, 1978; Prunell, 1983; Kefalas et al., 1988). Digestion pause sites and thus nucleosome boundaries can be mapped in relation to a DNA end by digestion of 5′ end labelled DNA.

In vivo, sequence effects will still drive the formation of nucleosomes at certain sequences, but higher order chromatin structure and the presence of other DNA binding proteins can also affect nucleosomal location. Nucleosomes can be mapped with nucleotide resolution *in vivo* using techniques such as ligation-mediated PCR (LM-PCR). LM-PCR works by

ligation of a linker sequence to the end of core DNA (e.g. from a MNase digest) and the amplification of core DNA by nested primers within the sequence and a primer corresponding to the linker DNA thus allowing the high resolution mapping of nucleosome boundaries.

However problems can arise *in vivo* as there will also be other DNA binding proteins present that may restrict access to DNA, potentially resulting in detection of false nucleosome boundaries. DNA at nucleosome boundaries does not interact as strongly with a nucleosome as DNA more centrally positioned on the nucleosome does (Weischet et al., 1978). DNA at the centre of a nucleosome binds the $(H3H4)_2$ tetramer whereas sequences at the nucleosome boundary bind H2A/H2B (Luger et al., 1997). If a nuclease is not fully halted at a nucleosome boundary nucleosome positions calculated may be weak or incorrect. Use of a combination of DNA cutting methods such as DNaseI and MNase can help overcome these problems, and will also even out any DNA sequence preferences of the different nucleases.

1.3.1 Nucleosome Structure Over the Ovine β-Lactoglobulin Gene.

The nucleosome structure over 10kb of the Ovine β-Lactoglobulin (BLG) gene and promoter has been mapped both *in vivo* and *in vitro* (Boa, 1999; Gencheva and Allan, 2005). The *in vivo* map compared the nucleosome structure in lactating mammary gland, the tissue BLG is expressed in, to the structure in liver; a non-expressing tissue. Many aspects of the two maps were similar, but they displayed some important differences. In liver, a regular nucleosome array is present throughout the gene and its promoter. In mammary tissue the array is retained up to the promoter, but the nucleosome structure becomes disrupted over the gene and few positioned nucleosomes are detected (illustrated in Figures 1.4 and Figure 3.1). The

terminal 3′ nucleosome in the mammary promoter array is actually a pair of two positioned nucleosomes. As only one nucleosome can actually sit over this DNA region in any cell, the fact that two alternatives are present in the active gene but not the inactive may point to a nucleosome structure-facilitated switch in the promoter.

The technique used to map the *in vivo* nucleosome positions does not map at high resolution. In order to be able to interpret the functional significance of the observed change in chromatin structure in the BLG promoter, information regarding the precise position of nucleosomes with respect to the DNA sequence would be invaluable.

An *in vitro* nucleosome positioning map of the BLG promoter and gene was produced using monomer extension (Boa, 1999; Gencheva and Allan, 2005). The region corresponding to the nucleosome pair mapped *in vivo* at the proximal promoter displayed a strong nucleosome positioning sequence, placing a nucleosome dyad at –183bp with respect to the transcription start site and a second strong overlapping site with a dyad at –224bp. The position at –183 is given the name nucleosome A and is represented in yellow in all figures in this thesis. The position at –224bp, is named nucleosome B, and likewise is represented in brown (Figures 3.2 and Figure 1.4). Nucleosome A sits almost exactly between two binding sites for the transcription factor Stat5, whereas nucleosome B lies over the upstream Stat5 binding site leaving only the downstream site external to the nucleosome.

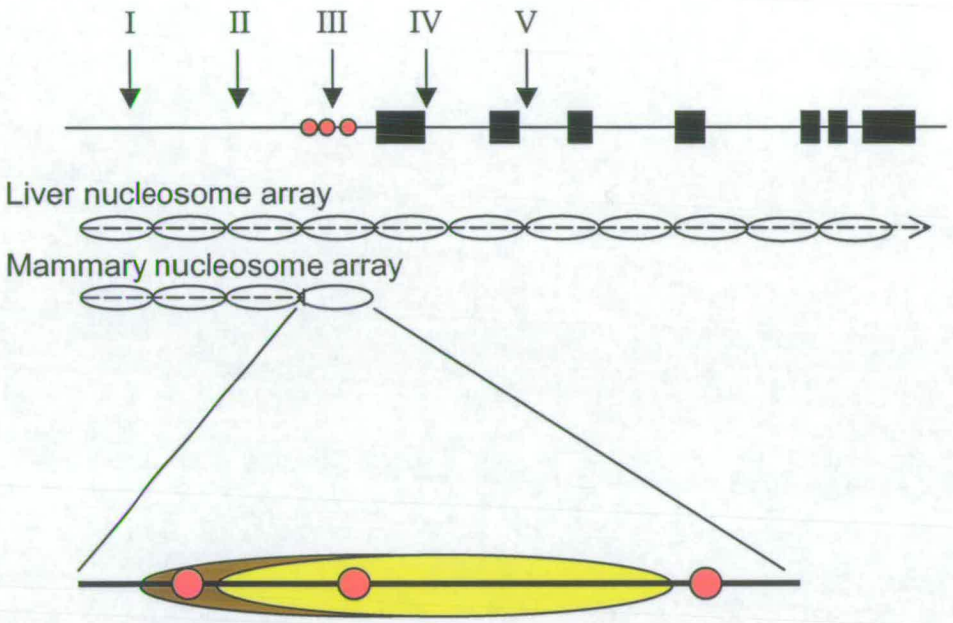


Figure 1.4 The BLG Gene.

The BLG gene contains 7 exons, marked by black boxes. DNaseI hypersensitive sites are marked by arrows and roman numerals. HSIII coincides with three Stat5 binding sites (red circles) and the point where the *in vivo* mammary nucleosome array running through the promoter terminates (indicated below the gene). Part of the region covered by HSIII has been enlarged below the array. Two *in vitro* mapped nucleosome positions A (yellow) and B (brown) differentially cover the Stat5 binding sites.

1.4 Milk Proteins.

Ever since nuclear proteins from lactating mammary gland were first observed binding to a milk promoter (Lubon and Hennighausen, 1987) the search has been on to define the factors responsible for the high level tissue specific and temporal regulation of milk proteins. Stat5 was first discovered as a mammary specific DNA binding protein, milk protein binding factor (MPBF) (Watson et al., 1991) or mammary gland factor (MGF) (Schmitt-Ney et al., 1991). Further characterisation revealed these to be members of the signal transducer and activator of transcription (Stat) family of transcription factors (Burdon et al., 1994a; Wakao et al., 1994) and the protein was renamed Stat5. Later it was realised there were two forms of Stat5, Stat5a and Stat5b (Liu et al., 1995). Stat5 regulates expression of milk proteins via prolactin induced activation of the Janus kinase (Jak) / Stat pathway (Heim, 1999; Barahmand-Pour et al., 1998; Darnell et al., 1994; Brelje et al., 2002).

1.4.1 The Proximal Promoters.

A conserved binding site for Stat5 is found in the region around –90 to –100 bp from the transcription start site in the promoters of many milk protein genes. This proximal site is responsible for the hormone responsiveness of β-casein (Wakao et al., 1994), BLG (Burdon et al., 1994a; Burdon et al., 1994b; Demmer et al., 1995) and αs1-casein (Pierre et al., 1994). Introduction of such a site into the murine α-lactalbumin proximal promoter introduces this property (Soulier et al., 1999). There is often more than one Stat5 binding site in the proximal promoter region (Vilotte and Soulier, 1992; Jolivet et al., 1992), in BLG these are located at –278, –210 and –93 bp from the transcription start site (Watson et al., 1991). In αs1-casein and β-casein these factors have been shown to act synergistically with a second set of Stat5 binding sites at a site further from transcription start, the distal promoter.

1.4.2 The Distal Promoters.

Many milk protein genes also contain a second set of Stat5 binding sites at a site further from transcription start (Millot et al., 2001; Pierre et al., 1994; Li and Rosen, 1995). These act cooperatively with the proximal Stat5 sites and other factors to produce maximal expression of the genes (Pierre et al., 1994). No functionally tested distal Stat5 binding sites have been reported in BLG, but there are Stat5 consensus binding sites at –2336 and –4014bp from the transcription start site. Half consensus sites or near-consensus Stat5 binding sites are spread throughout the promoter. Resection of the BLG promoter first to 3150bp and then to 400bp of 5′ flanking sequence leads to two reductions in the BLG promoter activity in CHO-K3 cells suggesting there are other factors responsible for the BLG gene activity in these regions (Demmer et al., 1995).

1.5 The BLG Gene.

BLG is the major milk whey protein in ruminants. It is not found in humans or rodents, but a transgenic mouse model exhibits a nearly identical expression profile to sheep (Whitelaw et al., 1992). BLG, as with most milk proteins, is expressed at high levels and has been exploited to produce pharmaceutical drugs such as factor IX (Schnieke et al., 1997) and α1-antitrypsin (Archibald et al., 1990) in sheep milk using the BLG promoter to drive expression specifically in the mammary gland. BLG is a member of the lipocalin family, and binds fatty acids. It may have a role in the transport of these and retinol (vitamin A) (Kontopodis et al., 2002).

1.5.1 DNase I Hypersensitive Sites in BLG.

There are five DNaseI hypersensitive sites (HS) in the BLG promoter (Figure 1.4); each has a distinctive temporal pattern and is only found in mammary tissue (Whitelaw and Webster, 1998; Whitelaw et al., 1992). HSI, located –1200 bp from the transcription start site, is detected from mid pregnancy, but is predominantly detected during lactation. The occurrence of HSII, located -800bp from transcription start, is erratic and weak. HSIII lies over the sequence from –300 to +100bp relative to the transcription start site, this includes the region of DNA where the change in the *in vivo* nucleosome structure occurs (Figure 1.4 and Figure 3.1). This is also where the Stat5 binding sites are located (Figure 1.4 and Figure 3.3). The extent of digestion at HSIII directly reflects the level of BLG expression. HSIII is detected weakly in virgin mammary gland, increasing in strength through late pregnancy to a peak during lactation before disappearing during involution. BLG is expressed at low levels in virgin mammary gland and expression increases from mid pregnancy through lactation until weaning and the subsequent involution of the mammary gland. HSIII is present in the correct pattern in mice transgenic for BLG, including those where the 5′ flanking sequence has been reduced to 406bp. This 406bp proximal promoter is also able to direct correct tissue specific and temporally regulated expression of BLG in transgenic mice (Webster et al., 1995; Whitelaw et al., 1992).

HSIV and V are present up to the first upregulation of milk protein gene expression at mid pregnancy, and may reflect a poised chromatin structure of the gene in mammary tissue. HSIV and V are 700bp and 1500bp downstream of HSIII respectively (Whitelaw et al., 1992; Whitelaw and Webster, 1998).

The HS sites over the sheep and goat BLG gene differ, HSIV is present during lactation in goat tissue but not in sheep. Sheep BLG is expressed at a higher level than goat but the promoter sequence is 96% similar, suggesting that the chromatin structure has a role in regulation of BLG. No HS sites are detected in liver (Pena et al., 1998).

1.5.2 BLG Promoter Expression Studies.

The role of the three Stat5 binding sites in the transcriptional activity of the proximal promoter has been studied in various promoter reporter studies. The proximal BLG promoter contains three binding sites for Stat5 only one of which (StM) fully obeys the recognised Stat5 consensus sequence (Figure 1.5). Transient transfection studies in CHO-K3 cells containing the long form rabbit prolactin receptor revealed that only StM is responsible for BLG promoter activity (Demmer et al., 1995). However in studies of transgenic mice and in the mammary epithelial cell line HC11, all three Stat5 binding sites are required for maximum activity and appear to act cooperatively (Burdon et al., 1994a; Burdon et al., 1994b). The major difference between the CHO-K3 and the other two systems is the mammary gland developmental background; the gene may be primed for activation ready to respond to lactation stimulus. Such a state could take the form of a preset chromatin structure over the gene as suggested by the presence of HSIV and HSV before BLG expression is upregulated. In transient transfections such as those in CHO-K3 cells the gene is not organised into chromatin, but the promoter-reporter constructs will be organised into a chromatin structure in both the transgenic mice and the stably transfected HC11 cells, neither of which contain endogenous BLG. HC11 cells express transfected BLG at a much lower level than transgenic mammary gland, this may reflect that the transgene has gone through mammary gland development in mouse, but not

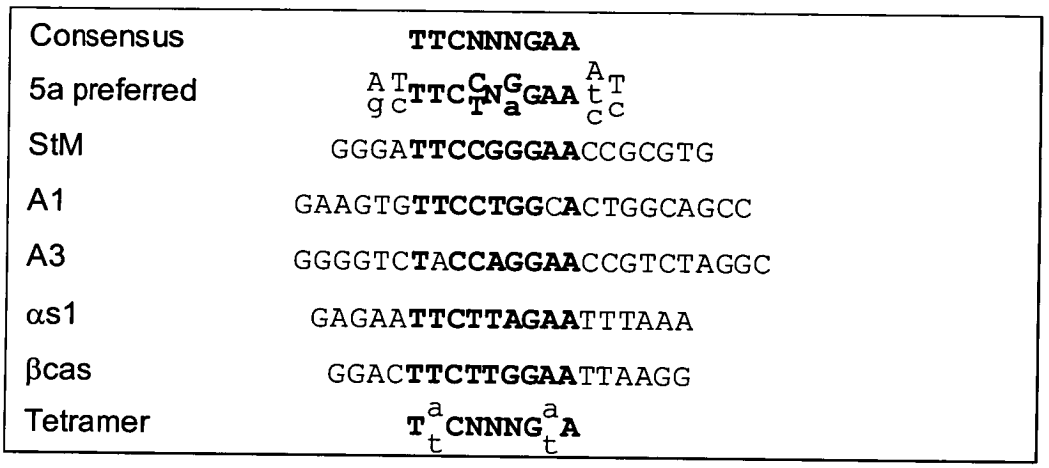


Figure 1.5 Stat5 Binding Sites.

Consensus binding sites for Stat5, the Stat5 consensus is shown in bold. StM, A1 and A3 are from the sheep BLG promoter, αs1 is the rabbit αS1 casein proximal Stat5 binding site and βcas is the rat β-casein proximal site. The consensus sequence shown for tetramer binding is for a second site, separated by 6bp from a strong site. The tetramer consensus is less well defined, as a full consensus is not necessarily required for binding, the sequence represented here has been demonstrated to encourage binding as a tetramer (Soldaini et al., 2000).

in HC11 cells. Developmental specific variation in chromatin structure may explain the different activities seen between CHO, HC11 and mouse assays for the proximal Stat5 sites.

High level expression and tissue and temporal specificity is conferred in transgenic mice by a minimum proximal promoter sequence of 406bp (Whitelaw et al., 1992; Webster et al., 1995). However in CHO-K3 cells expression from the 406bp promoter is much reduced compared to constructs containing an extended 5' sequence, up to 4.2kb (Demmer et al., 1995). Removal of the StM site from these longer promoters also abolished prolactin induction but this is not reproducible in mammary epithelial cells (Burdon et al., 1994a; Burdon et al., 1994b). In the light of the properties of other milk protein promoters, it is possible that a cooperative interaction between distal sequences and the proximal promoter might play a role in generating the maximal BLG expression. No analysis of the longer promoters has been carried out either in the HC11 cell line or mice and no HS sites coincide with the distal Stat5 binding sites.

1.6 Stat5.

BLG, like other milk proteins is activated in response to prolactin stimulation of the Jak / Stat pathway. Stat proteins are a family of cytoplasmic transcription factors. They were discovered initially through studies of interferon signaling (Darnell et al., 1994), but are activated by a wide range of cytokines and growth factors, see (Ihle, 1996), and (Darnell, 1997) for reviews. Stat genes have been identified in *Drosophila* (Yan et al., 1996; Hou et al., 1996) and in *Dictostelium*, (Kawata et al., 1997) but not in yeast indicating an appearance early in the evolution of eukaryotes.

1.6.1 Complex Regulation of Stat Factors.

Seven members of the Stat family have been identified so far in mammals, participating in many roles in the regulation of a whole host of cellular processes. These are numbered Stat1 to Stat6, including the two forms of Stat5, 5a and 5b. Stat1 is required for the interferon (IFN) pathway and Stat1 knockout mice have no innate immune response. Stat2 also is required for regulation of IFN-α (but not IFN-γ). Stat2 knockout mice are early embryonic lethal, as are Stat3 knockouts. Stat3 is activated by amongst other factors epidermal growth factor (EGF) and interleukin-6 (IL-6), and has a role in mammary gland involution. A large range of factors can activate Stats1, 3 and 5, but their actions are specific to cell type. Stat4 knockout mice are deficient in T helper 1 cells, and Stat6 knockout mice are deficient in T helper 2 cells. Stat4 is activated by IL-12 and IFN-α and Stat6 is activated by IL-4 and IL-13 (Horvath and Darnell, 1997; Darnell, 1997; Philp et al., 1996).

Stat proteins are found in pairs in mouse and human. Stat5 and Stat3 lie on mouse chromosome 11 and are 24% similar. Stats 1 and 4 lie on mouse chromosome 1 and are 50% similar and Stats 2 and 6 lie on chromosome 10 and are 19% similar. The low similarity suggests that the duplication causing these pairs occurred early in evolution of the gene. Stat5a and 5b are 96% similar in their peptide sequences suggesting this was a more recent duplication (Copeland et al., 1995).

1.6.2 Stat5a and Stat5b.

The two forms of Stat5, 5a and 5b were originally discovered during the cloning of murine Stat5 (Liu et al., 1995).

Stats 5a and 5b have partially redundant roles. Stat5a-/- mice develop normally but have mammary gland morphology defects and fail to lactate, Stat5b does not appear to be able to substitute for Stat5a in this situation (Liu et al., 1997) leading to the conclusion that Stat5a is the major Stat involved in lactogenesis. Stat5b -/- mice also show defects in lactation indicating that Stat5b plays a separate role in lactogenesis. Stat5b mice also show defects in reproduction, fat deposition, hair growth and sexually dimorphic liver gene expression. These characteristics can mostly be attributed to a loss of responsiveness to growth hormone (GH) (Heim, 1999; Udy et al., 1997). Stat5 binds to certain promoters in response to GH stimulation of the Jak/Stat pathway and the prolactin receptor has been shown to interact with the GH receptor (GHR) on ligand binding (Biener et al., 2003; Ooi et al., 1998; Woelfle et al., 2003). Stat5a/5b double mutants have a phenotype similar to GH or GHR null mice and are infertile (Teglund et al., 1998).

Stat5a and Stat5b play different roles in the regulation of milk protein production; Stat5b-/- mice produce WT levels of whey acidic protein (WAP) after 36 hours post partum (pp) but Stat5a-/- mice do not (Teglund et al., 1998). At 12 hours pp WAP expression is reduced by the same proportion in both knockouts compared to the WT. However neither appears to affect the sustained expression of β-casein individually suggesting that 5a and 5b act redundantly in regulation of this gene. In HC11 mammary epithelial cells, despite the presence of similar levels of Stat5a and Stat5b protein, Stat5a homodimers are the predominant DNA binding species (Cella et al., 1998; Teglund et al., 1998). The expression pattern of BLG in transgenic mice is more similar to β-casein than to WAP. The major upregulation of WAP expression occurs later in gestation than that of BLG or β-casein. BLG

expression is leaky in the mouse mammary gland and upregulation occurs earlier than β-casein upregulation (Harris et al., 1991). This leaky expression may reflect a tighter regulation of endogenous BLG in sheep, where the gene is located in the correct position within the genome.

1.6.3 Stat5 Mammary and Haematopoetic Roles.

Stat5 has also been shown to be activated by other signals including GH, IL-2, IL-3, IL-5, granulocyte-macrophage colony stimulating factor (GM-CSF), erythropoietin (EPO), thrombopoietin (TPO), and EGF (Ihle, 1996; Hoey and Schindler, 1998; Darnell, 1997; Sadowski et al., 1993). Thus Stat5 is also activated in cell types other than mammary, is involved in the regulation of other genes and is not mammary specific as was originally thought. Stat5 activity has been extensively studied in respect of its roles in erythropoiesis and thrombopoiesis; a double knockout of Stat5a and Stat5b shows non-fatal defects in these processes (Socolovsky et al., 2001; Teglund et al., 1998). IL-2 activates Stat5 as part of the regulation of the IL-2 receptor (Meyer et al., 1997; Rusterholz et al., 1999).

These facts would suggest that it is another factor, and not Stat5, that is involved in directing mammary specific expression of BLG. BLG is also expressed, albeit at low levels, in transgenic mouse mammary tissue where all three Stat5 binding sites have been mutated so that Stat5 is not able to bind (Burdon et al., 1994b). Nevertheless, mutations of the Stat5 binding sites have demonstrated an essential role for this transcription factor in the high expression level of BLG but not in tissue specificity (Burdon et al., 1994b; Demmer et al., 1995; Burdon et al., 1994a). Other factors may yet be shown to relevant in the proximal promoter, the sequence of which is sufficient to target BLG expression to the mammary gland. The mammary

specific signal does not necessarily have to be a transcription factor that can be detected binding to the DNA during lactation, it is possible that it may be encoded in a specific chromatin structure formed at an earlier developmental timepoint or by a combination of factors recruited by protein-protein interaction rather than by sequence specific DNA binding.

1.6.4 Mammary Gland Development.

It has long been known that insulin, hydrocortisone and prolactin act synergistically to produce milk proteins in the lactating mammary gland (Juergens et al., 1965). Both prolactin and Stat5 have been shown to be essential for mammary gland development. Stat5 is required for the formation of secretory alveoli, although prolactin null mice still form these structures, suggesting Stat5 does not always act through prolactin stimulation. Both prolactin receptors and Stat5 are required for milk protein gene expression (Miyoshi et al., 2001; Kelly et al., 2002). Over expression of Stat5 in mice leads to increased β-casein levels and a delay in apoptosis, this is enhanced if a constitutively active form of Stat5 is used. Such a protein also leads to increased tumour formation (Iavnilovitch et al., 2002). Misregulation of Stat5 may be a factor in breast cancers, Stat5a is active and nuclear localised in 76% of human breast cancers (Corlata et al., 2004). BRCA1 and BRCA2 have a role in the downregulation of Stat5, germline mutations of these genes increase susceptibility to breast cancer (Taniuchi et al., 1967). Another factor involved in mammary gland differentiation is MRG, a member of the FABP family. Mutation of this and the related MRGI have also been linked to breast cancer progression (Wang et al., 2003).

During normal regulation of the mammary gland a second Stat family member, Stat3, is involved in mammary involution at weaning.

Upregulation of Stat3 and downregulation of Stat5 negatively regulates mammary gland growth (Philp et al., 1996; Chapman et al., 1999).

1.7 Stat5 Action.

1.7.1 Stat Activation.

Several conserved domains exist between Stat proteins; these are described below and in Figure 1.6. The N-terminal domain is involved in protein-protein interactions, including those with nuclear cofactors such as CBP/p300 and between Stat dimers. It also has a role in negative regulation, as does the coiled coil domain (Strehlow and Schindler, 1998). The SH2 domain makes specific contacts with phospho-tyrosine of a second Stat, mediating dimerisation which allows the DNA binding domain to bind DNA (Barahmand-Pour et al., 1998; Shuai et al., 1994). The C-terminus contains the transactivation domain; phosphorylation of a serine in this area affects the transcription activity of Stat target genes, an effect which appears to be dependent on the target promoter (Beuvink et al., 2000; Park et al., 2001).

C-terminal truncated isoforms have been associated with the negative regulation of Stat5. These naturally occurring dominant negative forms of Stat5 dimerise and bind to DNA with a longer half life than full length Stat5, but they do not activate transcription as they lack the transactivation domain and thus their presence blocks binding sites for active Stat5 dimers (Wang et al., 1996; Moriggl et al., 1996). For reviews of Stat structure see (Ihle, 1996; Hoey and Schindler, 1998; Darnell, 1997; Sadowski et al., 1993; Meyer et al., 1997).

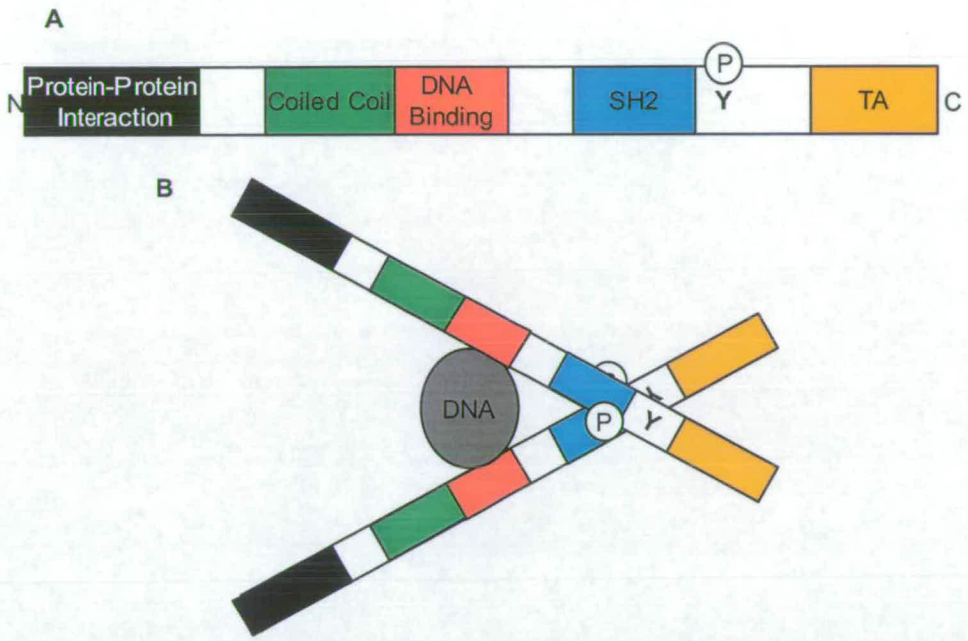


Figure 1.6. Domains and DNA Binding of Stat Proteins.

A. In this schematic representation of a Stat protein each conserved domain is represented by a coloured box. From the amino terminal (N) a black box marks the domain responsible for protein-protein interactions, the green box marks a coiled coil domain and the red box the DNA binding domain. The blue box marks the SH2 domain and the orange box the transcription activation (TA) domain at the carboxy terminal (C) of the protein. The conserved tyrosine which is phosphorylated on activation of Stat is marked by Y. **B.** Schematic representation of Stat as a dimer bound to DNA. Mutual interactions between the SH2 and the phosphorylated tyrosine (pY) cause two Stats to dimerise, DNA is bound between the two DNA binding domains.

Many groups are currently working on aspects of Stat-mediated signal transduction, as Stat proteins have roles in regulating many important cellular processes. Figure 1.7 presents a schematic representation of the general activation pathway for Stat proteins. Actual pathways are much more complex, as any one Stat can be activated and modified by many different factors and lead to regulation of numerous genes. The pathway, in the detail shown in Figure 1.7 and described below, is relevant to all Stat proteins.

Activation of the Jak/Stat pathway is initiated when an activator molecule, usually a cytokine or a hormone, binds to its receptor on the cell surface. A ligand-induced change in the receptor leads to auto-phosphorylation of Jak kinases associated with the intracellular domain of the receptor. These in turn phosphorylate a tyrosine residue on the receptor, which is recognized and bound by the Src homology 2 (SH2) domain of a latent Stat molecule in the cytoplasm. Latent Stat proteins were originally thought to exist as monomers, but recent evidence shows that latent Stats are also present as dimers in a conformation that is not translocated to the nucleus and is unable to bind DNA (Schroeder et al., 2004; Soler-Lopez et al., 2004). Stats recruited to activated receptors are themselves tyrosine phosphorylated by Jak and dissociate from the receptor as dimers, joined by mutual interactions between their SH2 domains and the conserved phospho-tyrosine residue (Y694 in Stat5a and Y699 in Stat5b) of their partner.

Stats form homo-dimers (e.g. 5a/5a or 5b/5b) and some can form hetero-dimers (e.g. 5a/5b); Stat5 forms both. Activated Stat dimers are translocated

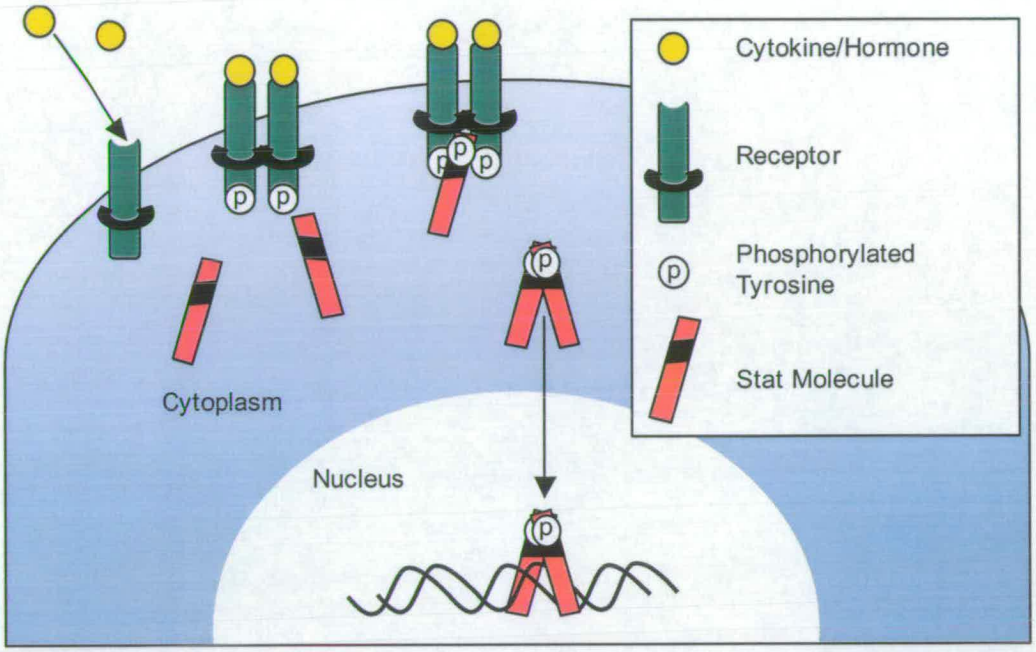


Figure 1.7. Jak / Stat Activation Pathway.

A schematic representation of the general Jak / Stat activation pathway. A cytokine or hormone binds to receptors in the cell surface resulting in receptor phosphorylation by Jak kinases associated with the receptor. A latent Stat protein is recruited to the receptor by interactions of the SH2 domain (marked as a black band on Stat proteins) with the phosphorylated tyrosine and the Stat is itself phosphorylated resulting in dimerisation. This Stat dimer is recruited to the nucleus where it binds DNA.

to the nucleus by an unknown mechanism, where they bind DNA and contribute to the activation of target genes. Stat proteins can interact with and be modified by other proteins at any stage in this process and such interactions are only beginning to be understood. Of course Stat5 binding to DNA is not just a matter of binding to Stat5 sites, as eukaryotic DNA is organised into chromatin and this may render Stat5 binding sites inaccessible. There are currently no reports in the literature describing how Stat5 binds within a chromatin context.

1.8 Stat5 Regulation: Interactions with Proteins.

Stat proteins show an activity specific to cell type and stimulus. As the preferred DNA binding sites are more or less identical for the majority of Stat proteins, it is unlikely that specificity is conferred via DNA binding. Specificity must therefore be conferred by other mechanisms.

1.8.1 Receptors and Jaks.

Stat specificity can be conferred by specific interactions with receptors. For example IFN activation of Stat5 has been shown to lead to the tyrosine phosphorylation and activation of different Stat5 isoforms in different cell types (Meinke et al., 1996). Latent forms of many Stat proteins are present in mammary gland, but only Stat5 is activated in response to prolactin. Two tyrosine residues in the C-terminal of the long form prolactin receptor have been identified that when phosphorylated will induce binding and activation of only Stat5a and Stat5b. The remaining Stats that cannot associate with the receptor retain their latent conformation (Mayr et al., 1998). A group of feedback inhibitors, the suppressors of cytokine signalling (SOCS) proteins, inhibit Stat proteins by binding to Jaks thus preventing the binding and activation of Stat proteins. SOCS are upregulated in response to prolactin

receptor activation in non mammary tissues during lactation, and in mammary gland after forced weaning (Tam et al., 2001). The selective activation of different Stat proteins and inhibition by SOCS proteins represent two mechanisms by which interactions with receptors could influence the mammary specificity of Stat5. More roles are likely to emerge as receptors could influence which Stat dimers are formed and modify residues other than the conserved tyrosine (Y694/Y699) that is required for Stat activation.

1.8.2 Non Jak-Mediated Stat5 Activation.

Stat5 can be activated through pathways other than the Jak/Stat pathway. Flt3 has been shown to activate Stat5a without activating any Jaks as has Src which activates both forms of Stat5, but with alternative phosphorylation at the C-terminal resulting in nuclear translocation of Stat5b but not Stat5a (Zhang et al., 2000; Kazansky et al., 1999; Kabotyanski and Rosen, 2003). ERBB family kinases have been shown to mediate Stat5 activity in mammary gland differentiation, and to phosphorylate Stat5 on tyrosine residues other than the conserved Y694/Y699 (Long et al., 2003). The functional role of these differentially phosphorylated Stat5s is still unknown.

1.8.3 HAT Activity.

Stat5 recruits transcription coactivators such as CBP/p300 to the promoter, thus stimulating transcription (Pfitzner et al., 1998). CBP/p300 are part of a family of transcription activators that contain HAT activity and are involved in the coactivation of many genes. CBP/p300 interacts with another family of coactivators, the p160/SRC/NcoA family. Stat5a binds a member of this family, NCoA-1, by an interaction involving a motif of three amino acids in the Stat5 transactivation domain, these same amino acids are essential for the

transcriptional activity of Stat5 (Litterst et al., 2003). In the β-casein promoter this interaction requires the GR (Wyszomierski and Rosen, 2001) and glycosylation of threonine 92 in the N-terminus of Stat5a (Gewinner et al., 2004).

Stat5 also interacts with HDACs, the function of this interaction is not to deacetylate histone tails, but to deacetylate CBP/p300. Acetylation of CBP/p300 results in its inactivation (Rascle et al., 2003; Xu et al., 2003). Stat5 has also been implicated in the negative regulation of genes in a mechanism that does not require Stat5 binding to DNA (Luo and Yu-Lee, 2003).

1.8.4 NF-1.

Another factor that plays a major role in the regulation of milk proteins is NF-1; there are binding sites for NF-1 in the promoters of milk protein genes, often situated close to those for Stat5 (Watson et al., 1991; Li and Rosen, 1995). NF-1B but not NF-1A acts synergistically with Stat5 and GR to activate transcription of WAP (Mukhopadhyay et al., 2001). Maximal induction of milk protein promoters in cell culture requires the addition of a combination of dexamethasone, insulin and prolactin. Dexamethasone binds and activates GR, and GR has been shown to interact with Stat5 *in vivo* (Cella et al., 1998). In the MMTV GR binding causes a change in nucleosome structure that allows NF1 binding. Maximal induction of milk protein expression probably requires the binding of Stat5, NF1 and GR to promoters.

1.8.5 SP-1.

Another general transcription factor, SP-1, has been demonstrated to interact with Stat5. NF-1, SP-1 and CBP/p300 have been shown to interact with specific transcription factors in a synergistic fashion in other promoters

(Aigueperse et al., 2001; Martino et al., 2001). SP-1 has been shown to be able to bind DNA organised into a nucleosome, the binding affinity appears to be related to the position of an SP-1 binding site on the nucleosome (Li et al., 1994). SP1 has not been shown to be essential for high-level expression of a milk protein gene, but could play a role in the establishment of an active transcription complex.

1.8.6 Transcription Factor Binding Sites in the Proximal BLG Promoter.

There are transcription factor binding sites for a number of factors in the proximal promoter of BLG. These include NF-1, SP-1 and GR, which have all been shown to interactact with Stat5 as well as YY1 and other sites such as NF-kappaB and C/EBP alpha (Figure 1.8).

1.8.7 Stat5 and Chromatin.

Stat5 has been shown to recruit CBP/p300 HAT activity and HDAC activity, but that these do not modify histones (Pfitzner et al., 1998; Xu et al., 2003; Rascle et al., 2003). Recently, acetylation of histone H3 and H4 was shown to be decreased on Stat5 activation of a Stat5 responsive promoter at the region of Stat5 binding. The authors suggest this correlates with chromatin remodelling at this locus resulting in the movement of a nucleosome (Rascle and Lees, 2003). Stat1 has been shown to be able to mediate expression from a chromatin template, this requires L724 and phosphorylation of S727 in the transactivation domain (Zakharova et al., 2003).

1.9 Stat5 DNA Binding.

The DNA binding domain is found in the centre of a Stat protein. It has a structure similar to an immunoglobulin fold, although there is little sequence identity to other DNA binding molecules containing this motif, such as

Figure 1.8 Transcription Factor Binding Sites on the BLG Promoter.

A search for transcription factor binding sites on the sequence nAB used in bandshifts was carried out using Transcription Element Search Engine on the WWW (TESS) of the Transfac database (http://www.cbil.upenn.edu/tess) (Schug and Overton, 1997). Selected transcription factor binding sites are shown.

KEY	
GR/PR	Boxed
NF1	*******
SP1	Underlined Black
YY1	Underlined Grey.
Stat5	**Red type Bold**
Stat5 half sites	Red type

NF -KB. Stats can only bind to DNA as an active dimer; interactions between the SH2 and DNA-binding domains mediate this (Chen et al., 1998). Stats bind at a palindromic motif, called the gamma activating sequence (GAS) consensus sequence after its original discovery in the IFNγ promoter, $TTC(N)_3GAA$. Variations in N may contribute further to the specificity of Stat/DNA interactions (Seidel et al., 1995; Schindler et al., 1995). Stat5a homodimers prefer to bind to (A/g)(T/c)TTC(C/T)N(G/a)GAA(A/tc)(T/c) and Stat5b homodimers to (A/tg)(T/A)TTC(C/T)(T/cag)(G/a)GAA(T/A)(T/ca) (the GAS sequence is underlined), but can bind as either homo- or hetero-dimers of the two forms of Stat5. These two sequences are very similar so it is unlikely that there is a difference in DNA binding site selection for the two Stat5 isoforms (Soldaini et al., 2000).

1.9.1 Structure of Stat Dimer Bound to DNA.

When Stat proteins bind to form dimers, they form a shape like pliers as suggested in Figure 1.6. Mutual interactions between the SH2 domains and the phosphorylated tyrosine form a hinge, allowing DNA to sit between the DNA binding domains. DNA is bent by 40° away from the dimer interface and is almost completely enclosed by Stat binding, this would suggest Stat binding is not compatible with nucleosome binding. The N-terminal domains, forming the handles of the pliers, are placed far apart from each other (Figure 1.6). The image in Figure 1.6 is a schematic representation of how a Stat dimer bound to DNA would look, based on the crystal structures of Stat1 and Stat3 dimers bound to DNA solved by (Chen et al., 1998) and (Becker et al., 1998) respectively. These structures do not resolve the N-terminal domains.

1.9.2 Stat Tetramerisation.

Stats-1, -3, -4 and-5 have been shown to tetramerise on tandemly linked binding motifs (Vinkemeier et al., 1996; John et al., 1999; Meyer et al., 1997; Zhang and Darnell, 2001; Yamamoto et al., 2002). It has been suggested that Stat5 may be able to bind to non-consensus sites as a tetramer of two dimers. For example, a change of TTC to TAC in the GAS consensus will lower Stat binding as a dimer, but allows tetramer binding, resulting in formation of only tetramer complexes on a probe containing this sequence (Verdier et al., 1998; Soldaini et al., 2000). The extra stability conferred by the tetramerisation interaction allows binding at these non-consensus sites. The A1 and A3 Stat5 binding sites in the BLG promoter are non-consensus sites which contain either a TAC or the compliment of this, GTA (Figure 1.5). Stat5a makes tetramerisation interactions much more efficiently than Stat5b does (Verdier et al., 1998; John et al., 1999). This may represent another mechanism for the different actions of Stat5a and Stat5b.

The N-terminal domain, which mediates tetramerisation of Stat4, has been crystallised separately. The structure revealed a conserved tryptophan residue (37 in Stat5) in a flexible 'hook' which probably makes contacts with glutamic acid 66 (in Stat4) to form a tetramer (Vinkemeier et al., 1998; John et al., 1999). This structure was used to develop mutation studies, where removal of the conserved tryptophan blocks Stat tetramer formation, indicating that this area is responsible for tetramerisation. Mutation of this tryptophan to an alanine (W37A) also abolished tetramerisation interactions (John et al., 1999). DNA binding site and Stat tetramerisation deficient mutants have been used to show the functional importance of: Stat3 tetramerisation in activation of the α2 macroglobulin gene (Zhang and Darnell, 2001): Stat4 tetramerisation in IL-12 induction of the perforin gene

promoter (Yamamoto et al., 2002) and of Stat5 tetramerisation in the induction of the hepatic serine protease inhibitor 2.1 and in IL-2 activation of the IL-2 receptor α (IL-2Rα) chain (Bergad et al., 1995; John et al., 1996). The Stat5 tetramerisation mutant, W37A, can only activate the IL-2Rα promoter if the Stat5 binding sites are mutated from a non-consensus and a consensus pair to two consensus sites allowing two Stat5s to bind as dimers (John et al., 1999) this suggests that it is the binding of multiple Stat5 dimers which activates this promoter.

1.10 Stat5 Tetramerisation on the BLG Promoter?

To allow tetramer formation, Stat5 binding sites must usually be situated close together on the DNA. The optimum spacing for Stat5 appears to be 6bp; sites spaced further apart will also allow the interaction but sites spaced closer than 5bp do not. Binding sites must also be arranged in a tandem repeat (Vinkemeier et al., 1996; Soldaini et al., 2000). The N-terminus appears to contain a degree of flexibility, and is able to make interactions from a range of DNA binding site spacing. It may be possible that bringing the sites together by means other than by their adjacent location on a DNA sequence in cis will have the same effect. Winding of DNA round a nucleosome has been shown to bring together regulatory elements in a manner that enhances gene activation, Section 1.2.2.

It is possible that the strongly positioned nucleosome at –183bp on the proximal promoter acts to bring together the StM and A1 Stat5 binding sites in such a way as to facilitate a tetramerisation interaction enhancing BLG promoter activity (Figure 1.9). The major form of Stat5 that is involved in lactogenesis, Stat5a, is able to form tetramers much more efficiently than Stat5b can. This suggests a role for tetramerisation in facilitating Stat5a

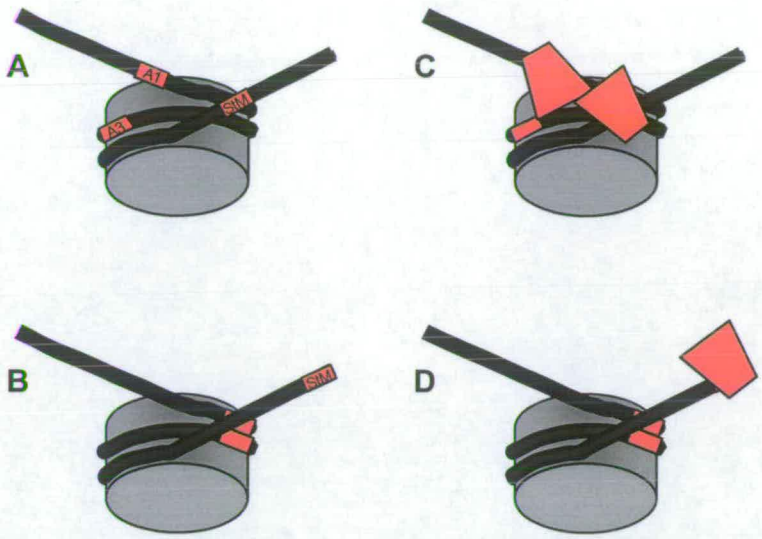


Figure 1.9. The Tetramerisation Model.

A schematic representation of the location of Stat5 binding sites on nucleosome positions A and B. The structures in parts **A** and **C** represent nucleosome A and those in parts **B** and **D** represent nucleosome B. Parts A and B represent the Stat5 binding positions on DNA and parts C and D represent how Stat5 might bind to DNA. The wide end of the wedge that represents a bound Stat5 dimer is the N-terminus, where dimer-dimer interactions take place.

specific mammary Stat5 action. Furthermore, the A1 binding site in BLG is non consensus, containing GCA instead of GGA. Such sequences have been show to encourage binding of Stat5 as a tetramer (Soldaini et al., 2000). The availability and relative location of each Stat5 binding site will be dramatically different between the two alternative nucleosome positions at –183 (Figure 1.9 A and C) or –224 bp (Figure 1.9 B and D) from the transcription start site. It is not known whether Stat5 is able to bind DNA in or near to a nucleosome. Either nucleosome A at –183bp or nucleosome B at –224bp could position the Stat5 binding sites in a favourable manner for BLG gene activation. Nucleosome B places the consensus StM binding site further from the nucleosome (Figure 1.9 B), which may allow preferential Stat5 binding to this site. On the other hand nucleosome A contains two Stat5 binding sites external to the core DNA (Figure 1.9 A) an arrangement that would potentially allow a tetramerisation interaction between the two bound Stat5s if the orientation is correct. Alternatively such an orientation could have the reverse effect and the binding of Stat5 to one site could block binding to the other.

It is clear that the regulation of BLG is affected by both the binding of Stat5 and the chromatin structure, indeed the chromatin structure may influence Stat5 binding and *vice versa.* In this study the positions of nucleosomes reconstituted onto a fragment of the BLG promoter, nAB, are defined. The ability of Stat5 to bind to sites close to and within these nucleosome is studied and a potential role of tetramerisation in stabilising this interaction is investigated.

2 MATERIALS AND METHODS.

2.1 Reagents, Stock Solutions and Buffers.

All chemicals are supplied by BDH and stored at room temperature unless otherwise stated. Solutions are in dH_2O unless otherwise indicated. All autoclaved reagents were sterilised by autoclaving at 15 pounds per square inch for 15 minutes at 121°C.

A/NT/L: Solutions A, NT and L at ratio of 2:3:5. Antiprotease and antiphosphatase solutions added prior to use. Those added were: Dithiothreitol (DTT) to 0.4mM, β–Mercaptoethanol (BME) to 5.6mM, 1% antiprotease mix from Sigma and 1% antiphosphatase mix from Sigma.

A/NT: Solutions A and NT at 1:1 ratio containing antiprotease and antiphosphatase as above.

Acrylamide: 40% solution at acrylamide:bis ratios of 19:1 or 37.5:1 from National Diagnostics.

Ammonium Persulfate: Catalyst for acrylamide polymerisation. 25% aqueous solution stored at +4°C. Fresh stock made weekly.

Ampicillin: (Sigma) stock at 4mg/ml. Final concentration 50µg/ml.

Beta-mercaptoethanol (BME): (Sigma) stock at 14.3M.

Buffer C: 20mM Hepes pH7.9, 25% glycerol, 0.42M NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA. Inhibitors PMSF to 0.5mM and BME to 0.5mM.

Coomassie Gel Stain: 50% methanol, 10% acetic acid, 0.1% Coomassie.

Destain (Coomassie stained gels): 10% acetic acid, 10% isopropanol.

DNA Size Markers: 1kb markers and massruler markers from Fermentas. 10bp markers from Invitrogen.

DMSO: Dimethyl sulphoxide (Sigma).

DTT: (Melford labs) 100mM stock stored at –20°C.

EDTA: (Ethylene diaminetetra-acetic acid disodium salt) stock at 0.5M, adjusted to pH8 with NaOH.

1 X ExoIII Buffer: 10mM Tris.HCl pH8, 10mM NaCl, 0.75mM $MgCl_2$, 1mM BME. Stored at –20°C as 10 X stock solution.

5 X Ficoll Loading Buffer: 5 X TBE, 17.5% (w/v) Ficoll 400 (Amersham) with bromophenol blue and xylene cyanol.

Formamide gel loading buffer: 95% formamide, 20mM EDTA, Bromophenol blue and xylene cyanol.

Gentamycin: Stock at 10mg/ml, final concentration 7μg/ml.

2 X GRB 1.6% (w/v) Ficoll 400, 2mM Hepes (Sigma) pH7.9, 1mM EDTA pH8.0. Solution is dissolved at 37°C and DTT added to 0.1mM. Store at –20°C.

IPTG: (Melford labs) Isopropyl-1-thio-β-D-galactoside, stock at 100mM.

Kanamycin: (Melford labs) Stock at 10mg/ml final concentration 50μg/ml.

KCl: Stock solution at 1M.

LB plates: 7g agar in 500ml LB medium. Agar is dissolved and sterilised by autoclaving. X-gal, IPTG and antibiotics are added just prior to pouring 10ml cooled agar to sterile petri dishes.

LB medium: 10mg/L Bacto Tryptone, 5mg/L Yeast extract, 10mg/L NaCl. Autoclave to sterilise.

NaAc: Sodium acetate, 3M stock solution pH 5.0 with acetic acid.

NaCl: Stock solution at 5M.

PBS 10 X Phosphate buffered saline. Stock solution was made up by addition of 10 tablets (from Oxoid) per 100ml and autoclaving to sterilise.

PMSF: Phenylmethanesulphonyl fluoride (Sigma). Stock of 250mM, dissolved in isopropanol and stored at 37°C.

10 X PNK buffer: 500mM Tris.HCl pH7.6, 10mM $MgCl_2$, 10mM BME.

Poly(dI.dC)•Poly(dI.dC): (Poly(deoxyinosinic-deoxy-cytidylic) acid) (Sigma) non-specific competitor DNA routinely used in bandshifts.

Ponceau S: 0.1% in 5% acetic acid.

Phosphate buffer: 0.02M sodium phosphate, 0.5M NaCl.

Protein size markers: Either Precision Protein Standards from BIORAD or Benchmark His Tagged protein standards from Invitrogen.

Radioactive nucleotides: [α^{32}P] dCTP (3000ci/mmol) and [γ^{32}P] ATP (3000ci/mmol) from Amersham.

SDS: Sodium dodecyl sulphate, 10% stock solution.

SOC medium: From Invitrogen. 0.5% yeast extract, 2% tryptone, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose.

Solution A: 0.6M Sucrose, 120mM KCl, 15mM NaCl, 0.3mM spermidine, 2mM. spermine, 28mM BME, 4mM EDTA, 2mM EGTA, 2mM DTT, 0.2% triton X-100, 10mM Tris.HCl pH 7.9.

Solution L: 15mM NaCl, 10mM Tris.HCl pH 7.9.

Solution NT: 10mM NaCl, 0.1% NP40, 10mM Tris.HCl pH 7.9.

Solution I: 15mM Tris-HCl pH8, 10mM EDTA, 100µg/ml RNase A (for minipreps).

Solution II: 0.2M NaOH, 1% SDS (for minipreps).

Solution III: 3M potassium acetate (for minipreps).

TAE 1 X working solution: 40mM Tris acetate, 2mM EDTA. Stock at 50 X.

20 X TBE: 1M Tris 1M boric acid, 20mM EDTA for native polyacrylamide gels, use at 0.25 to 1 X.

10 X TBE: 0.5M Tris pH8, 0.5M Boric acid, 10mM EDTA. Use at 1 X for denaturing polyacrylamide gels.

10 X TBS: 1.5M NaCl 250mM Tris.pH 7.5. Autoclave to sterilise.

TBST: 1 X TBS containing 0.05% Tween 20.

20 X TE: 2mM EDTA and 200mM Tris pH8 use at 1 X.

TEMED: (N,N,N',N'-Tetramethyl-ethylenediamine) Cross linking reagent for polyacrylamide gels.

TEP: 1 X TE containing PMSF at 10mM.

TEP_{800}: 1 X TEP containing 0.8M NaCl.

TEP_{10}: 1 X TEP containing 10mM NaCl.

Tetracycline: Stock solution 12mg/ml in 70% ethanol, final concentration 12μg/ml.

1M Tris Buffer: Tris (hydroxylmethyl) methylamine. Stock at 1M aqueous solution. pH 7.5 → 8 with HCl. Autoclave to sterilise.

Triton X-100: (Iso-Octylphenoxypolyethoxyethanol) from Sigma.

Tween 20: (polyoxyethylene(20)-sorbitan monolaurate) from Sigma.

Xgal: (5-bromo-4-chloro-3-indoyl-β-D-galactoside). Stock 20mg/ml in dimethyl formamide, stored at –20°C protected from light.

2.2 Preparation of DNA.

2.2.1 Transforming Competent Cells.

Plasmids were introduced into either ultracompetent (XL-10 Gold), or supercompetent (XL-1 Blue) cells (both from Stratagene) by chemical transformation. Approximately 10ng of a ligation reaction, or <1ng plasmid DNA was added to aliquots of thawed competent cells and incubated on ice for 10-30 minutes. Cells were heat shocked at 42°C for 30 seconds then allowed to recover on ice for at least two minutes. After addition of 500µl prewarmed SOC medium, cells were incubated at 37°C with shaking at 200rpm for one hour. In order to isolate individual clones, transformation reactions were plated onto LB-agar plates containing a selective antibiotic, resistance for which is encoded in the plasmid DNA, usually this was ampicillin. 100µg/ml X-gal and 40µg/ml IPTG were added if blue white screening (section 2.3.2) was to be used to select colonies. For straightforward plasmid transformations 5µl transformation mix was diluted in 200µl SOC and for ligation reactions half of the total mix was plated. Antibiotic concentrations routinely used are 100µg/ml ampicillin, or 7µg/ml gentamicin.

2.2.2 Small Scale Preparation of DNA.

Around 10µg DNA was routinely produced from a 4ml culture using the following miniprep protocols. Individual colonies were picked from plates using a sterile pipette tip, inoculated into 4ml LB containing selective antibiotics and incubated at 37°C with rotation at 200rpm overnight until they had reached stationary phase (approximately 16 hours). Cells were pelleted then resuspended in 300µl Solution I. 300µl Solution II was added to lyse cells, then the genomic DNA and protein was precipitated by dropwise addition of 300µl Solution III and centrifugation at 16,100g for ten minutes.

Supernatant containing the plasmid DNA was transferred to a tube containing 0.8ml ice-cold isopropanol to precipitate the plasmid DNA. Precipitated DNA was pelleted by centrifugation at 16,100g for 15 minutes, and washed twice with 0.5ml 70% EtOH. Air-dried plasmid DNA was resuspended in dH_2O or TE. Alternatively minipreps were carried out using Promega Wizard Plus SV Miniprep kit that makes use of spin columns to speed up the process.

2.2.3 Large Scale Preparation of DNA.

A larger culture volume (100-400ml) can be utilised to produce greater amounts of DNA. Routinely ~250μg DNA was produced using the QIAGEN midi prep kit.

2.3 DNA Purification.

Many applications require the purification of DNA. This was carried out by a number of protocols.

2.3.1 Gel Purification.

Specific DNA fragments were isolated from a reaction, for example a restriction digest product, by separation on an agarose gel. 10 X loading buffer (60% glycerol, 0.25% Orange G, 10 X TAE) was added to the sample before loading onto an agarose gel (between 0.8 and 2% (w/v) depending on fragment size), in 1 X TAE. Ethidium bromide in the buffer (at 250μg/L) intercalates with the DNA and fluoresces under long range UV light (304nm). Appropriate bands were identified by comparison of migration against 1kb DNA ladder (Fermentas), and excised from the gel using a clean scalpel. DNA was easily removed from the gel slice by use of the QiaexII or Qiaquick gel extraction kits (both from Qiagen). The gel slice was dissolved

in buffer QG at 50°C. This buffer contains chaotropic salts, which disrupt the gel structure. DNA from dissolved gel slices was bound to silica particles and agarose and residual salts were washed out using an ethanol buffer. The DNA was eluted into 30µl TE.

2.3.2 Phenol / Chloroform Extraction.

Nucleic acids were isolated from protein, for example after nuclease digestion, by phenol / chloroform extraction. Hydrophobic protein groups are soluble in phenol, whereas hydrophilic DNA is soluble in aqueous solution. Chloroform assists the process by denaturing proteins and therefore aiding dissociation from DNA (Current Protocols in Molecular Biology). The protein precipitates in the layer between the aqueous and phenol phases. This method is useful when total DNA is to be recovered.

DNA/protein mixtures are mixed by vortexing (fragments less than 1kb only) at a ratio of 1:1 with phenol (that has been equilibrated to pH7.6 by repeat extraction with Tris buffer). Microcentrifugation at 12,000g for 5 minutes separates the phases. The top aqueous phase was removed to a fresh tube and the phenol extraction repeated once. An equal volume of chloroform:isoamyl alcohol (24:1) was added to remove residual traces of phenol and the aqueous phase removed as described for phenol extraction. DNA was precipitated using ethanol (section 2.3.3).

2.3.3 Ethanol Precipitation.

DNA in a large volume of aqueous solution was precipitated to a pellet then resuspended into a smaller volume at higher concentration. DNA in solution is made to 300mM NaAc by addition of 3M NaAc, pH5. At least 2 X volume prechilled 100% ethanol is added to this. Incubation at either –70°C for one

hour or –20°C overnight precipitated the DNA out of solution. Samples were spun for 15 minutes at 12,000g. The supernatant was discarded and washed with prechilled 70% ethanol, this step removes excess salt, the ethanol concentration is sufficient to prevent DNA coming back into solution. After a further spin for 5 minutes, the 70% ethanol was removed as completely as possible. The pellet was then washed in an ether/ethanol mix (50:50) and either air dried or vacuum desiccated to remove ethanol before resuspension in dH_2O or TE. All residual ethanol was removed as this can affect downstream reactions. Over drying of pellets also adversely affected their ability to go back into solution, particularly with larger DNA fragments.

2.3.4 Calculating DNA Concentration.

The concentration of DNA in solution was calculated from absorbance at 260nm. For a solution containing double stranded DNA in a 1cm cell, one OD_{260} unit indicates a DNA concentration of 50μg/ml. Often the amount of DNA available was not sufficient to calculate the concentration in this way. In this case a known volume of the sample was run on an agarose gel alongside a known quantity of DNA and an estimate taken from the relative intensity of fluorescence of the known and unknown bands.

2.4 Manipulation of DNA.

2.4.1 Restriction Enzyme Digestion.

DNA was digested by various restriction enzymes as per the manufacturers instructions (Roche or New England Biolabs). Digestion was verified by agarose gel electrophoresis. Restriction maps were generated using NEBCutter V2.0 (http://tools.neb.com/NEBcutter2/index.php) (Vincze et al., 2003).

2.4.2 Cloning.

Sequences were cloned into plasmid vectors by ligation of DNA ends from complementary restriction sites on the vector and insert. Where this was not possible, ends were blunted with either T4 polymerase or the Klenow fragment of DNA polymerase following manufacturers instructions (NEB). Ligation reactions were carried out with the Rapid DNA ligation kit (Roche). If a fragment is to be cloned using a single enzyme, the vector DNA ends were first dephosphorylated with calf intestinal alkaline phosphatase (CIAP, (NEB)) to prevent religation of the vector. Phosphate groups on the insert DNA allow the backbone to be joined. Ligated DNA was transformed into *E. coli* to select for correct insertions. Many commercial cloning vectors, such as pBluescript, use disruption of a lacZ gene to screen for insertions. Colonies with a vector containing the insert remain white and religated vector colonies appear blue and on agar plates contain X-gal and IPTG. This is called blue white screening. Restriction enzyme digestion and sequencing was used to verify the orientation and correct size of the insert. Ligated circular DNA does not transform into *E.coli* as well as supercoiled plasmid DNA so more must be added to a transformation reaction. It is important to note however that too much DNA will have an adverse affect on the cells.

Recombinant baculovirus DNA (Bacmid DNA) was produced by transposition of the Stat5 gene, cloned into a pFastbac expression vector, into the baculovirus genome in DH10bac cells which contain a helper plasmid encoding a transposase. Transfected cells are grown with tetracycline (10μg/ml) gentamycin (7μg/ml) and kanamycin (50μg/ml) to select for colonies containing recombinant viral DNA. Insertion of the Stat5 gene destroys the lacZ gene so individual recombinant colonies can be recognised and picked for small-scale DNA preparation. High molecular weight DNA

(>100kb) is purified using the protocol described in (2.2.2) and transfected into insect cells (2.6.3). pFastbac vectors and DH10Bac cells are part of the bac-to-bac baculovirus expression system from Invitrogen. For more details on cloning into pFast vectors refer to Chapter 4.3.2.

2.4.3 Polymerase Chain Reaction.

Polymerase Chain Reactions (PCR) amplify the region of DNA between two oligonucleotide primers. The primers were cyclically denatured, allowed to anneal to the template sequence, and the primer extended along the length of the template by polymerase. Multiple cycles of this amplify a specific region of DNA. The sequence synthesised in the first round can act as a template in the second and so on, doubling the product with each round of amplification. PCR reactions were carried out generally as follows: Primers were chosen from the flanking regions of DNA and synthesised by MWG Biotech. The melting temperature (Tm) of primers was calculated using the formula:

$$Tm = 2 \times (A+T) + 4 \times (C+G).$$

A value 4°C lower than this was used as the annealing temperature. PCR reactions were catalysed using Taq polymerase and buffers from Roche, except for reactions where products were expected to be greater than 1kb, in which case the Expand High Fidelity PCR system (Roche) was used. This system includes Tgo polymerase as well as Taq polymerase. Tgo polymerase has proofreading properties resulting in a more accurate PCR product. dNTPs were added at a final volume of 2mM and template at 1ng/µl. Reactions were carried out in a volume of 50µl in a geneAmp PCR system 9700 (Applied Biosystems). After an initial denaturing step of 95°C for 5 minutes, 35 cycles of:

95°C for 20 seconds denaturing step

54°C for 30 seconds annealing step

72°C for 30 seconds elongation step

Followed by further elongation at 72°C for 5 minutes to ensure all products are full length. Primers were removed by the Qiagen PCR cleanup kit. The product was further purified by gel extraction (section 2.3.1).

2.4.4 Site Directed Mutagenesis.

Point mutations can be carried out using a modified PCR. Complementary primers to both strands of the DNA containing at most a couple of base pair changes from the template were designed. Changed bases were located in the middle of the primer, and certainly not near the 3′ end, as this must anneal tightly to the template in order for elongation to occur. Five cycles were carried out at a lower annealing temperature calculated by 2°C less than the normal primer annealling temperature for every mismatched A or T and 4°C less for every mismatched C or G. The remaining cycles were carried at the normal annealing temperature as there should be sufficient product containing the mutation to use as a template. PCR reactions were either carried out as normal, or to create mutations in a plasmid vector, were carried out using pfu turbo polymerase (Stratagene) on a supercoiled plasmid, this polymerase catalyses extention of the whole plasmid. A single sequence was used as the primer, but as two complimentary oligonucleotides containing the mutation resulting in two strands of the plasmid containing the mutation. Methylated DNA from the original template plasmid sequence is digested by *Dpn*I, this enzyme will not digest the PCR product. The resulting DNA was transformed into competent cells and mutation verified by sequencing. This method was adapted from the Quikchange kit, developed by Stratagene.

2.4.5 Sequencing.

All DNA sequencing was performed by The Sequencing Service (School of Life Sciences, University of Dundee, Scotland http://www.dnaseq.co.uk) using Applied Biosystems Big-Dye Ver 3.1 chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer.

2.5 Radioactive Labelling of DNA.

All radioactive work and disposal was carried out as per local and national guidelines on the use of radioactive isotopes. Radioactive nucleotides were from Amersham.

2.5.1 5′ End Labelling.

PCR products or oligonucleotides were 5′ endlabelled using T4 Polynucleotide kinase (PNK) (Amersham). T4 PNK catalyses the transfer of a phosphate group from the γ position of ATP to the 5′ hydroxyl terminus of polynucleotides. In this method the probe is labelled at both ends of double stranded DNA. Reactions were carried out essentially as detailed below.

50ng DNA in 1 X PNK buffer and 30 units of T4 PNK were incubated with 0.74 MBq [γ^{32}P] ATP for one hour at 37°C, then heat inactivated for 10-20 minutes at 68°C. it was found that addition of DTT to 10mM increased the efficiency of the reaction. Unincorporated [γ^{32}P] ATP was removed by passing the reaction through a ProbeQuant G50 (for larger DNA products) or G25 Micro Column (labelled oligonucleotides) (both from Amersham).

DNA ladders were labelled by a similar reaction. Where DNA ends are already phosporylated, phosphate groups are first removed using 20U CIAP in supplied buffer for 30 minutes at 37°C. The CIAP was heat inactivated at

75°C for 15 minutes before the end labelling reaction was carried out, there was no need to purify DNA between reactions.

2.5.2 α^{32}P Labelling.

For uniform labelling throughout the probe 0.74 - 1.11 MBq [α^{32}P] dCTP was included in an otherwise normal PCR reaction. Following the PCR, unincorporated α^{32}P was removed by passing the reaction through a G50 microcolumn. The reaction was gel extracted from a 2% agarose gel to purify the correct sized fragment and to remove primers.

2.6 Production of Stat5.

2.6.1 Baculovirus Expression System.

Recombinant Stat5 was produced using a baculovirus expression vector system (BEVS). This system was chosen as Stat5 produced in this way has been used previously in bandshift experiments (Soldaini et al., 2000; John et al., 1999). The advantages of using the BEVS are that proteins produced in insect cells are modified in a similar manner to mammalian cells (Smith et al., 1983), and are expressed at relatively high levels, which can be greater than 50% of the total cell protein (Maeda et al., 1985; Miyamoto et al., 1985). Insect cells also adapt well to growth in suspension culture allowing cultures to be easily scaled up. *Spodoptera frugiperda* (SF9) cells were used and infected with recombinant *Autographa californica* Multiple Nuclear Polyhedrosis Virus (*Ac*MNPV). Recombinant Stat5 expression was driven by the polyhedrin promoter, which in the wild type virus makes up the virus coat particles observed as occlusion bodies in SF9 cells infected with the wild type virus.

2.6.2 SF9 Cells.

SF9 cells were typically maintained at 27°C in suspension culture using a spinner system at densities between $3x10^5$ and $2x10^6$ cells/ml in serum free medium (SF900II from Invitrogen) containing penicillin at 100U/ml and streptomycin at 100µg/ml. Over time, cultures maintained in spinner culture and serum free medium lost viability, growth rates slowed and eventually cells began to die. Fresh cultures were seeded regularly and discarded after approximately 30 passages. A fresh culture was used for each protein expression experiment.

Resuscitated cells were seeded at $1x10^7$ cells per T75 flask, and the medium was replaced when cells were attached, typically after 1-2 hours. After 4-5 days cells were split into five T75 flasks. By this time cells are >100% confluent but remain viable. SF9 cells are not tolerant to trypsinisation and when grown in serum free medium attach firmly to the flask. At such high density cells begin to detach and are easily harvested. Sufficient cells were harvested from 3-5 over-confluent T75 flasks to start a 100ml spinner culture at $1x10^6$ cells/ml. Cells took approximately two weeks, or three passages, to become adapted to spinner culture conditions during which time the stirring speed was gradually increased from 50 to 80 rpm. When cells were fully adapted growth rates increase so that cultures were split two or three times in a week and cells no longer formed aggregates.

Samples were counted daily using a haematocytometer to follow growth rates (Figure 2.1). Only cells from the logarithmic phase were used. Regular growth curves were produced to show logarithmic growth, an example is shown in Figure 2.1. The number of cells in one of the nine major squares were counted over at least four squares, more squares if the cell count is

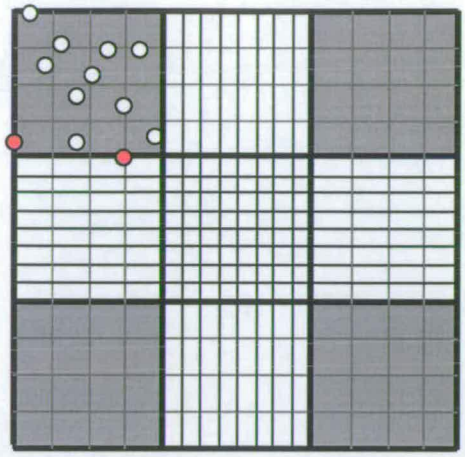
lower than $1x10^6$ cells per ml. Cells overlapping the bottom or left edge of minor squares are not counted to avoid counting cells twice. The volume in each of the major squares is $0.1mm^3$, or $1x10^4$ ml therefore the average number of cells in each major square $x10^4$ gives the number of cells per ml. Exclusion of trypan blue was used to show cell viability where required. Damaged cells have been shown to take up the blue stain but viable cells exclude it.

Mid log phase cells were frozen at $1x10^7$cells/ml in a 50:50 mix of fresh and conditioned medium containing 7.5% DMSO. Cells were wrapped in tissue to insulate them from sudden temperature changes and kept at –20°C for 5 hours; –80°C for 24 hours then stored long term at –150°C. Aliquots were taken after transfer to –150°C and resuscitated to check cell viability.

2.6.3 Transfecting SF9 Cells.

$9x10^5$ mid log phase cells were seeded per well of a six well plate and allowed to attach for one hour. During this time 1µg bacmid DNA was mixed with 100µl Grace's medium (Invitrogen) and in a separate tube 6µl Cellfectin reagent (Invitrogen) also was mixed with 100µl Grace's medium. These were combined and left for 45 minutes at room temperature to allow formation of DNA-lipid complexes. Cells were washed with 2ml Grace's medium, 0.8ml Grace's medium was added to the lipid/DNA complex and added to the cells. Non transfected wells were included as a control. The plate was incubated at 27°C for five hours after which time the transfection mix was removed and replaced with SF900II containing antibiotics. Transfected cells were incubated in a bag containing moist towels (a humidified incubator). After 72 hours, infected cells began to lyse and show late signs of infection. Cells did not detach in great numbers until after 96 hours post infection.

A.



B.

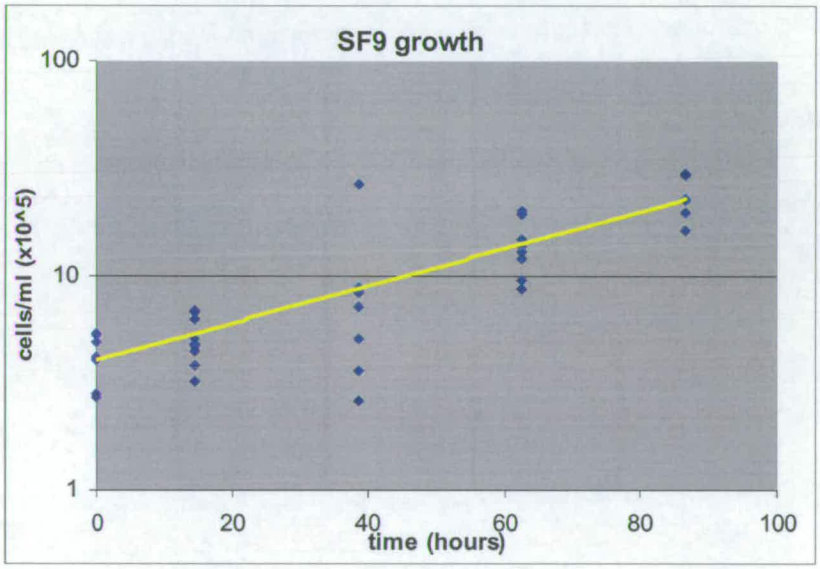


Figure 2.1 SF9 Growth.

SF9 cells were counted at various time points using a haematocytometer (**A**). At least four samples per time point were counted, in the major squares marked in grey. Only cells lying completely within the square (white circles) or over the upper or right hand boundary were counted, cells lying over the other boundaries were not counted (red circles). (**B**) The best-fit line when plotted as a log graph is linear, indicating exponential growth.

When cells reached late stages of infection, medium was removed and clarified by centrifugation at 500g to remove cell debris. This P1 viral stock was then amplified by infection of 2x 10^6 cells in a 6 well plate at an estimated multiple of infection (MOI) of 0.1. Medium was collected and clarified after 48 hours to give the P2 viral stock. A further amplification step in a T75 flask was carried out to give a higher volume virus stock (P3).

2.6.4 Virus Titre.

The Viral titre in the P2 or P3 stock was calculated using the BacPAK baculovirus Rapid Titer Kit from BD Biosciences. 6 x10^4 SF9 cells per well in 96 well plates were infected in quadruplicate with 50µl viral stock at dilutions between 10^{-3} and 10^{-5}. After incubation in a humidified incubator for 43 hours cells were fixed and wells blocked with 3.5% goat serum in PBS containing 0.05% Tween 20 (PBST). The cells were then incubated with a mouse antibody against an AcMNPV envelope glycoprotein, gp64, at a dilution of 1:200 in goat serum in PBST. This antibody will detect and bind to infected cells. A secondary goat anti-mouse-IgG antibody containing a HRP conjugated enzyme was used at 1:250. Foci of infected cells were visualised as blue stained cells by addition of blue peroxidase substrate. The average number of foci was counted over the three dilutions and the virus titer calculated by the following equation:

Virus titre (pfu/ml) = average number of foci per well X dilution factor X40 X2

Multiplication by 40 normalises for inoculum volume to give titer in focus forming units (ffu) per ml. Multiplication by 2 is a conversion factor

empirically calculated by BD Biosciences to give the titre in pfu/ml. All reagents were supplied by BD Biosciences.

2.6.5 Virus Infection.

For large scale production of rStat5, 7.5 x10^6 mid log phase SF9 cells were seeded in a T75 flask and infected at an MOI of 10. These were incubated in a humidified incubator at 27°C for 72 hours before harvesting.

2.6.6 Harvesting Recombinant Stat5.

Medium was collected from SF9 cells before cells were harvested after viral infection. Cells were washed twice with phosphate buffer to remove traces of medium. A lysis buffer is added which consists of phosphate buffer containing antiprotease and antiphosphatase mixes from Sigma plus 0.05 times volume Insect Popculture reagent (Novagen). This contains a detergent that breaks up the cell membranes during a 15 minute incubation at 4°C. Lysed cells were frozen at –20°C then spun at 3,200g in an Eppendorf benchtop centrifuge to remove cell debris and DNA from the lysate before loading onto a column for purification of the recombinant protein. Over expression of rStat5 can easily be seen when whole cell extract is analysed by SDS PAGE (Figure 4.8).

2.6.7 Purification of 6-His Tagged Stat5.

Recombinant Stat5 produced in the baculovirus expression system was purified using Ni^{2+} chelating affinity for a 6-His tag engineered into the N terminus of the protein. 1ml columns packed with chelating sepharose (Amersham) were charged with a solution of Ni^{2+}. Histidine residues form complexes with the metal ions and proteins with these groups available are retained in the column (Porath et al., 1975). A group of 6 histidine residues

positioned together has a high affinity for the column. The presence of imidazole in the binding and elution buffers competes the histidine association with the nickel ions.

Various imidazole concentrations for the binding and elution buffers were tested, 30mM imidazole for the binding buffer and 300mM imidazole for the elution buffer was found to be best for optimum purification of the 6-His tagged Stat5. Columns were equilibrated with the binding buffer before a cell extract in binding buffer was loaded onto the column, the flow through was collected for analysis of protein binding. The column was washed with 10ml binding buffer to minimise binding of other proteins and aliquots collected for analysis. After addition of 5ml elution buffer the eluate was collected in 0.5ml fractions. The 6-His tagged Stat5 normally eluted from the column in the second and third fractions. A sample was run on SDS polyacrylamide gel electrophoresis (PAGE) to verify purification and the condition of rStat5 produced (Figure 4.8). Yield was between 3 and 6.3μg rStat5 per 10^6 cells.

The 6-His tag of the Stat5W37A was designed to be removed by cleavage with AcTEV protease (Invitrogen). Digests were carried out at room temperature in buffer supplied with the enzyme. After digestion samples were stored at –20°C before analysis on protein gels.

2.6.8 Preparation of Nuclear Extracts.

Nuclear extracts were prepared from lactating mammary gland, a tissue in which the BLG gene is active, and from liver as a negative control. Tissue from lactating sheep was used, as sequences used in these studies are from the ovine BLG gene. Mid-lactation (11 days post partum) mouse tissue was

also used, as this was readily available. Extracts were prepared by a method based on (Dignam et al., 1983). Tissue was ground to a fine powder using a pestle and mortar in liquid nitrogen and cells were resuspended at 100ml per 20mg tissue in A/NT/L (at a ratio of 2:3:5). Cells were lysed by several strokes of a dounce homogeniser, filtered through four layers of Miracloth (Calbiochem) to remove any connective tissue debris and washed once in 10ml A/NT/L, and once in A/NT. The washed pellet was collected by centrifugation at 1,200g for ten minutes and resuspended in buffer C containing inhibitors at 3ml per 10^9 cells. Nuclei were lysed by gentle agitation with a magnetic stirrer for 30 minutes at +4°C. The extract was cleared by centrifugation at 25,000g for 30 minutes. Protein concentrations measured by the BCA assay kit (section 2.7.3) were between 2 and 4 mg/ml.

2.7 Protein Analysis.

2.7.1 SDS PAGE.

Protein samples were analysed by SDS PAGE. Prepoured gels and buffers were bought from Invitrogen and gels run according to manufacturers protocols. NuPAGE 12% Bis Tris and Tris acetate gel systems were used. Total proteins were visualised by staining with Coomassie gel stain for one hour, then destaining with destain solution. Bands were visible after two hours destaining and further destaining in fresh solution overnight resulted in a clear background. 6-His tagged proteins were visualised by staining with Invitrogen's InVision™ His-tag in gel stain. Gels were fixed for one hour in a solution containing 40% ethanol and 10% acetic acid, then rinsed twice in dH_20 for 10 minutes. The gel was stained with the InVision™ solution for at least one hour, but preferably overnight before washing twice with phosphate buffer for ten minutes. An image was taken after this initial wash and after any subsequently required further washes. Gels were loaded

with Benchmark™ His-tagged protein standard (Invitrogen) which also served as a positive control for the stain. The maximum excitation wavelength for the InVision™ stain is 560nm and the maximum emission is 590nm. Images were taken using a fluor fx multi-imager from BIORAD, using the closest settings to these maxima (excitation at 532nm and a long pass filter of 555nm).

2.7.2 Western Blotting.

Samples were separated by SDS PAGE and transferred to Protran BA83 nitrocellulose membrane (from Schleicher and Schuell) using the NuPAGE system from Invitrogen following manufacturers instructions. Transfer of protein was verified by Ponceau S staining. Membranes were blocked overnight in a solution of TBS containing either 3% BSA, or 5% non-fat dried milk (Marvel). Incubations with primary antibodies and secondary (HRP-conjugated) antibodies were at dilutions recommended by suppliers in blocking buffer, plus 0.05% Tween 20 to minimise non-specific interactions. Bands were visualised by a chemiluminescence reaction using the PIERCE SuperSignal West Pico system and images captured onto Kodak film.

Densitometry analysis was carried out using Quantity One software (Molecular Dynamics), after films were converted into a digital format using the densitometry / x-ray film function of a Fluor S multi-imager (imager and software from BioRad). The signal from each band was measured over a duplicated area and normalised to background. The percentage of the signal over the total selected area was calculated for each band. Data was exported to Microsoft Excel and normalised to the lane loading using data from a Western blot with the loading control anti-SF9 cell line antibody.

2.7.3 Measuring Protein Concentration.

Protein concentration was measured using the BCA assay kit from PIERCE. 25µl unknown samples and BSA standards diluted in the same buffer as the samples, were incubated in a 96 well plate in triplicate with 200µl of the supplied working solution at 37°C for 30 minutes. Reduction of Cu^{2+} to Cu^{+} by protein in an alkaline solution is detected by bicinchoninic acid, and the resulting colour change to purple was measured at A_{560} after subtraction of a blank which was the dilution buffer in working reagent. Data were exported to Microsoft Excel and a BSA standard curve generated from which the protein concentration of the samples was calculated.

2.8 Bandshift Analysis.

Bandshift or Electrophoretic Mobility Shift Assays (EMSAs) are a sensitive and reliable tool used to detect DNA protein binding complexes, They have the ability to detect femtomolar quantities of protein binding to DNA (Current Protocols in Molecular Biology, volume 2). Protein and labelled probe were incubated together in the presence of a non-specific competitor DNA (Poly(dI.dC)•Poly(dI.dC)), to minimise non-specific interactions. Complexes were separated by migration through non-denaturing PAGE in TBE. Naked DNA migrates at a much faster rate than DNA associated with protein. DNA-protein stability is maintained by the 'cage effect' whereby the polyacrylamide gel physically keeps DNA and protein in close proximity so that if they were to dissociate the likelihood of their re-association is increased.

2.8.1 Non Denaturing Polyacrylamide Gels.

Bandshift reactions were separated by non-denaturing PAGE. Depending on the application these gels contain 4-6%(v/v) 37.5:1 polyacrylamide, in 0.25 to

1X TBE. Acrylamide/TBE mixtures are mixed with the crosslink reagent TEMED (1.25µl/ml) and the polymerisation catalyst APS (313µg/ml), the gel mix was poured into a the space between two glass plates thoroughly cleaned with 70% ethanol and dH_2O that were separated by 1mm plastic spacers. Gels were prerun for one hour. Wells were flushed with running buffer immediately before loading to remove any traces of non-polymerised acrylamide. Samples were either loaded in Ficoll loading buffer or for Stat5 bandshifts the reaction was loaded without loading buffer as the bandshift buffer contains sufficient Ficoll. After the gel had finished running, the plates were separated and the gel was vacuum dried at 80°C for one hour to Whatman DE81 anion exchange chromatography paper.

2.8.2 Preparation of Oligonucleotide Probes.

Oligonucleotide probes for bandshift reactions were created by annealing two complimentary oligonucleotides (synthesised by MWG biotech) corresponding to Stat5 binding sites on the BLG promoter. 50pmol of the sense strand was 5′ end-labelled as described above. A 6 fold excess of the unlabelled antisense oligonucleotide was added and the mixture brought to >95°C for 5 minutes then allowed to cool slowly to room temperature. 5X Ficoll loading buffer was added and the mixture loaded onto the minimal number of wells on a 10% polyacrylamide gel in 1X TBE. This was run at 200V constant until the bromophenol blue dye was approximately half way down the gel.

The glass plates were split open and loading buffer mixed with some ^{32}P labelled DNA spotted onto pieces of filter paper arranged strategically round the edges of the gel. The glass plate plus gel and spots was sealed into a plastic bag and exposed to a phosphor screen. After one hour the screen was

scanned using a fluor fx multi-imager (BioRad). A real size image was produced, and a printout of this was used to orientate the blue dots on the gel with the dots on the phosphorimage. The phosphorimage also revealed the location of the nucleosome position isomers thus allowing the appropriate bands to be isolated.

The double stranded oligonucleotide band was cut out and eluted in 200µl TE containing 50mM KCl_2 resulting in a final double stranded probe concentration of 250 fmol/µl.

2.8.3 Stat5 Bandshift Analysis.

Stat5 bandshifts were carried out using either 20-40µg of a mammary nuclear extract or 1-2µg purified recombinant Stat5 (rStat5) produced in the baculovirus expression system. This equates to between 11 and 22 pmoles of rStat5. Stat5 was incubated with 1µg Poly(dI-dC)•Poly(dI-dC) (a non specific competitor) in buffer GRB. This mix was incubated on ice in a total volume of 20µl for 15 minutes before the addition of 1µl radiolabelled probe (0.25 pmol, which normally contained between 50 and 100 cps) and any competitor DNA. This equates to a molar ratio of rStat5 to probe DNA of between 44:1 and 88:1.

Unlabelled competitor DNA competes with the labelled probe for Stat5 binding, and when added at a known molar excess was a useful tool for comparing the affinity of Stat5 for various DNA sequences. Reaction mixes were incubated for a further 15 minutes at room temperature (21°C). Samples were loaded straight onto a 5% non-denaturing polyacrylamide gel in 1X TBE, with 0.25X TBE running buffer. Gels were run at 200V constant for three hours. An empty lane was loaded with GRB containing

bromophenol blue and xylene cyanol to follow migration. In a 5% gel the fragment nAB migrates just ahead of the xylene cyanol.

2.8.4 Nucleosome Reconstitution by Salt Gradient Dialysis.

Nucleosomes were formed on a fragment of the BLG promoter, nAB, by the salt gradient dialysis method. Initial reconstitutions were carried out at ratios of 20, 30, 40, and 50:1 donor chromatin to DNA. These looked similar when run on a gel so subsequently only 20:1 was used, as even at this ratio, the majority of the probe is associated with protein. Donor chromatin used was chicken erythrocyte monomers, stripped of linker histones and other proteins. Only the four core histones are present.

TEP buffers were prechilled to +4°C and PMSF was added to 2.5mM to inhibit proteases. A dialysis sheet with 6000/8000Da exclusion limits was prepared by soaking in dH_2O treated with PMSF. For reconstitution at a ratio of 20:1, 4.95 µg chicken erythrocyte donor chromatin (at 825ng/µl from Dr J. Allan) and 250ng probe were incubated in TEP containing 1.2M NaCl for 30 minutes at 37°C. This was applied to the membrane of a microdialyser (Invitrogen) previously equilibrated with prechilled TEP containing 0.8M NaCl (TEP_{800}) and dialysed against 500ml TEP_{800} at 2ml/minute for 4 hours. The salt concentration was brought down further to 10mM NaCl by overnight dialysis against 1 litre of TEP_{10}.

2.8.5 Isolating Positioned Nucleosomes.

Nucleosomes positioned at different sites on a DNA fragment can be separated by their migration through a non-denaturing polyacrylamide gel. A nucleosome at the centre of the fragment is more retarded in its migration through the gel than a nucleosome sitting at the end of the same DNA

fragment (Pennings, 1997). I have made use of this behaviour to separate the different positioning isomers formed on nAB.

Reconstituted nucleosomes were separated on a 5% non-denaturing polyacrylamide gel in 1X TBE. Gels were prerun for at least one hour at +4°C and run at 60V constant overnight. Nucleosome positioning isomers were maintained at a low temperature to minimise nucleosome movement. Individual nucleosome positioning isomers were separated simply by cutting each band out of the gel as described for isolating the annealed oligonucleotide probes (section 2.8.2). Cut out bands were eluted into 200µl TEP_{10} containing carrier chromatin at 60ng/µl. Samples of the cut out bands are run alongside the original reconstitution mix on a non-denaturing polyacrylamide gel to verify the correct isolation of the positioning isomers.

2.9 Mapping Nucleosome Positions.

DNA associated with nucleosomes is protected from the action of nucleases. Digestion of unprotected DNA to identify nucleosome boundaries enables calculation of the location nucleosomes position at on a given DNA fragment. In this study positions of nucleosome on reconstituted DNA probes were mapped by two different enzymatic approaches. The methodology employed with each approach is quite different, thus the combination of data from the two approaches provides a better indication of where on the DNA nucleosomes are positioned.

2.9.1 Exonuclease III Mapping.

Fragment nAB of the BLG promoter, which includes the three Stat5 binding sites, was produced by PCR using primers designed to incorporate a *Pvu*II restriction site at the upstream end and a *Rsa*I restriction site at the

downstream end. See Figure 2.2 for primer sequences. The fragment was 5′ end labelled using T4 PNK and [γ^{32}P] ATP. The engineered restriction sites facilitated the removal of 8-9bp from the ends of the fragment resulting in a molecule labelled at only one end, and on only one strand. This fragment was reconstituted with nucleosomes from carrier chromatin as described previously (2.8.4).

Exonuclease III (Exo III) cuts away at the 3′ end of duplex DNA in a stepwise manner (Richardson et al., 1964). It shows some sequence specificity of C>A=T>G (Linxweiler and Horz, 1982) and thus a partial digest of DNA will show a unique pattern of pause sites. The presence of a positioned nucleosome causes ExoIII to pause at the nucleosome boundary. However, ExoIII will eventually cut into the nucleosome producing a characteristic 10bp periodicity in pause sites (Linxweiler and Horz, 1985; Riley and Weintraub, 1978; Kefalas et al., 1988).

NaF*Pvu*II	GGCTCTcAg*CTGTCCTTGTCTAAG
NbR*Rsa*I	GGCTGGt*aCCCCAGCCACGCGGT

Table. 2.1. Primers used to introduce *Pvu*II or *Rsa*I restriction enzyme sites into nAB.

Changed bases are shown in lower case. Restriction sites created are underlined and cleavage sites marked by*.

50ng DNA or 250ng isolated nucleosome isomers were incubated at 37°C for 15 minutes with 25 – 3,000 U/ml ExoIII (NEB) in 1X ExoIII buffer. One U of ExoIII is defined as the amount of enzyme required to produce 1nmol acid soluble nucleotide in 30 minutes at 37°C in a total reaction volume of 50µl. Reactions were stopped by adding EDTA to 15mM and SDS to 0.2% and by heating to 70°C for 20 minutes. Samples were phenol/chloroform extracted,

ethanol precipitated and resuspended in formamide gel buffer and single stranded DNA products were resolved on a denaturing gel.

Data were analysed using AIDA (Advanced Image Data Analysis) software. The gel was calibrated by the migration of known labelled DNA standards, which in this case was a 10bp ladder from Invitrogen. The position of the label at the 5′ end of the strand is known. From this each pause site for ExoIII was accurately mapped to its position on the fragment. By using DNA labelled at first one 5′ end then the other both boundaries of each nucleosome were mapped in this manner.

2.9.2 Denaturing Polyacrylamide Gel Electrophoresis.

Single stranded DNA lengths were determined by electrophoresis in a denaturing 8% polyacrylamide gel containing 1x TBE and 8M Urea at 50W for about three hours.. Prior to loading on a gel, DNA was denatured by heating to 95°C for five minutes then put straight onto ice before loading onto a gel prerun to 55°C. The gel apparatus was disassembled, prior treatment of one plate with Repel silane ensured the gel only ever stuck to one plate. Gels were fixed in a solution of a 10% acetic acid, 12% methanol solution and rinsed with dH_2O. Fixed gels were transferred on to 17Chr Whatman chromatography paper and vacuum using a Bio-Rad 583 gel dryer. Gels were visualised after an initial overnight exposure using a phosphorimager, then exposures of up to one month for further analysis.

2.9.3 Restriction Mapping.

The second enzymatic approach to mapping nucleosome positions used a combination of micrococcal nuclease (MNase) digestion to produce core

DNA, and subsequent restriction enzyme digestion. MNase pauses at nucleosome boundaries (Noll and Kornberg, 1977; Horz and Altenburger, 1981). Nucleosomes were reconstituted onto nAB that had been labelled uniformly by addition of [α^{32}P] dCTP into the PCR reaction used to produce the fragment. (See labelling DNA 2.5.2).

Core particle DNA was produced by digestion of unfractionated reconstitutes. 50µg/ml reconstitute in TEP_{10} containing 1mM $CaCl_2$ was digested with MNase (10 units/ml) for 20 minutes on ice followed by further trimming to the nucleosome boundaries at 21°C, at times individually calculated by timepoint digests. Isolated core particle DNA was analysed by denaturing polyacrylamide gel electrophoresis to check the quality of the 146bp band.

Core protected DNA was digested by unique cutting restriction enzymes. Products of these digestions were separated on a denaturing gel alongside 10bp ladder. An enzyme cleaving within the core DNA produces two products, the sizes of which add up to 146bp. The digest products from cleavage by two or more enzymes within a fragment facilitated mapping of nucleosome boundaries with respect to the known restriction sites.

3 CHARACTERISING NUCLEOSOMES POSITIONS ON nAB

Several DNaseI hypersensitive sites detected throughout the BLG promoter and gene mark areas of interesting chromatin structure (Whitelaw and Webster, 1998). Hypersensitive site III is of particular interest as its appearance is linked to the activity of the BLG promoter. HSIII sits at the proximal promoter, immediately before the transcription start site. The appearance of DNaseI hypersensitive sites represents a change in the accessibility of DNA to nuclease attack. This can be due to transcription factor interactions, or to the displacement or movement of a nucleosome. The presence of HSIII suggests a link between the chromatin structure of the promoter and the regulation of BLG expression. For a more in-depth description of the DNaseI hypersensitive sites in BLG refer to section (1.5.1).

3.1 Introduction: Nucleosome Positions on the BLG Promoter.

In order to study a role for chromatin in the regulation of the BLG gene, positions of nucleosomes over the BLG promoter have been studied both *in vivo* and *in vitro* by our lab (Boa, 1999; Gencheva and Allan, 2005). From this data a clearer picture of the role positioned nucleosomes may play emerges.

3.1.1 Nucleosome Positions *In Vivo.*

In vivo nucleosome binding positions have been mapped throughout the BLG gene and promoter in tissues that either express, or do not express BLG. A partial cuprous phenanthrolene digest of chromatin, combined with digestion at a unique restriction site was utilised to map nucleosome positions by indirect-end-labelling (Wu, 1980). Extracted DNAs were separated by agarose gel electrophoresis, transferred to nitrocellulose and probed with a sequence adjacent to the restriction enzyme site. Part of the resulting nucleosome map derived from this analysis is shown in Figure 3.1.

A major reason for studying BLG is that its expression is limited to the lactating mammary gland. Differentiation of the mammary gland takes place in the adult; hence tissue at different developmental stages is readily accessible. Results of the *in vivo* mapping experiments show a regular nucleosome array is present throughout both the promoter and the gene in liver (Figure 3.1 lanes 4 and 5) where the BLG gene is in its inactive state. A different structure is detected in the active tissue, where the nucleosome array is seen at the 5′ end of the promoter but not in the transcribed region (Figure 3.1 lanes 2 and 3). The last (most 3′) nucleosome remaining from the array through the promoter is detected over the area of DNase HSIII, which includes binding sites for the transcription factor Stat5. The indirect labelling indicated that there are two weak nucleosome positions at this site (marked by the upper set of (stick) arrows to the left of the gel in Figure 3.1). The nucleosome position 5′ to this is also a pair, but is much stronger (marked by the lower pair of (block) arrows in Figure 3.1). Importantly this data indicates that a nucleosome is located over the proximal promoter in the active state and constitutes an integral component of the HS itself.

The technique used to map nucleosome positions *in vivo* is relatively low resolution and lacks the accuracy to precisely define where on the DNA individual nucleosomes lie. Importantly, it does tell that nucleosomes form an array over a particular area of DNA, and provides an indication of where the array is disrupted. However the significance of this arrangement is not yet clear. Active genes are often found to have a more open and accessible chromatin structure (Felsenfeld et al., 1996). However the fact that the array ends at a DNaseI hypersensitive site, where there are two alternative nucleosome positions, may be significant and could suggest a potential

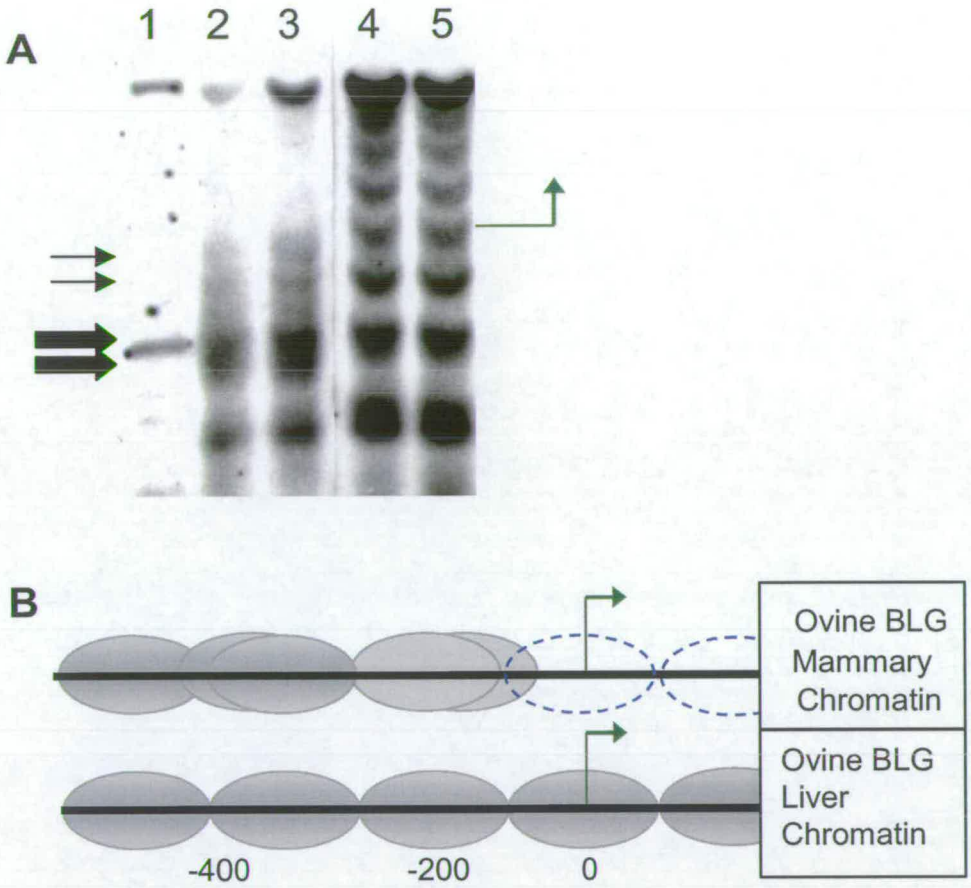


Figure 3.1 *In Vivo* Chromatin Structure over the BLG Gene.

A regular nucleosomal array is seen throughout both the promoter and transcribed region in liver, but in the active mammary gene this pattern is disrupted throughout the expressed region. The indirect end-labelling analysis is shown in **A.** Lanes 2 and 3 show the map of the *in vivo* nucleosome positions in sheep mammary gland. The two pairs of arrows on the left indicate the two pair of alternative nucleosome positions. Lanes 4 and 5 show the nucleosome map in liver. This data is represented in schematic form in **B,** showing the alternative nucleosome positions. Green arrows represent the transcription start site and grey ovals represent nucleosome positions. Lane 1 is genomic DNA digested with RsaI plus 1kb ladder. The relative strength of the nucleosome positions is represented by the shade of grey. Both these images were adapted from Figure 52 of (Boa, 1999).

mechanism for positioned nucleosomes in regulation of the BLG gene.

3.1.2 Nucleosome Positions *In Vitro.*

To complement the *in vivo* analysis, *in vitro* nucleosome binding positions, determined solely by the interactions of the histone core with the DNA sequence, were also mapped over the BLG gene and promoter using the monomer extension technique (Yenidunya et al., 1994; Boa, 1999; Gencheva and Allan, 2005). In this approach, nucleosomes are reconstituted onto the DNA of interest and digested with MNase to produce 146bp of core particle DNA protected by the nucleosome. The monomer DNA recovered from the core particles is then labelled and annealed to a single stranded version of the original template DNA and extended to a known restriction site. The sizes of the resulting products, measured on a denaturing polyacrylamide gel, identifies the boundary of the parent nucleosome. The band intensity also reflects the preference of nucleosomes to sit at certain points on the DNA.

Using this technique, which enables accurate mapping of nucleosome binding sites over a long stretch of DNA, several strong individual nucleosome positions were detected throughout the gene (Figure 3.2). Of particular note was a strong nucleosome positioned with its centre, or dyad, at –183 base pairs (bp) relative to the transcription start site, referred to hereafter as nucleosome A (Figure 3.2).

The BLG promoter harbours three binding sites for the transcription factor Stat5 (Watson et al., 1991), the outer two of which are located at –93 and –278

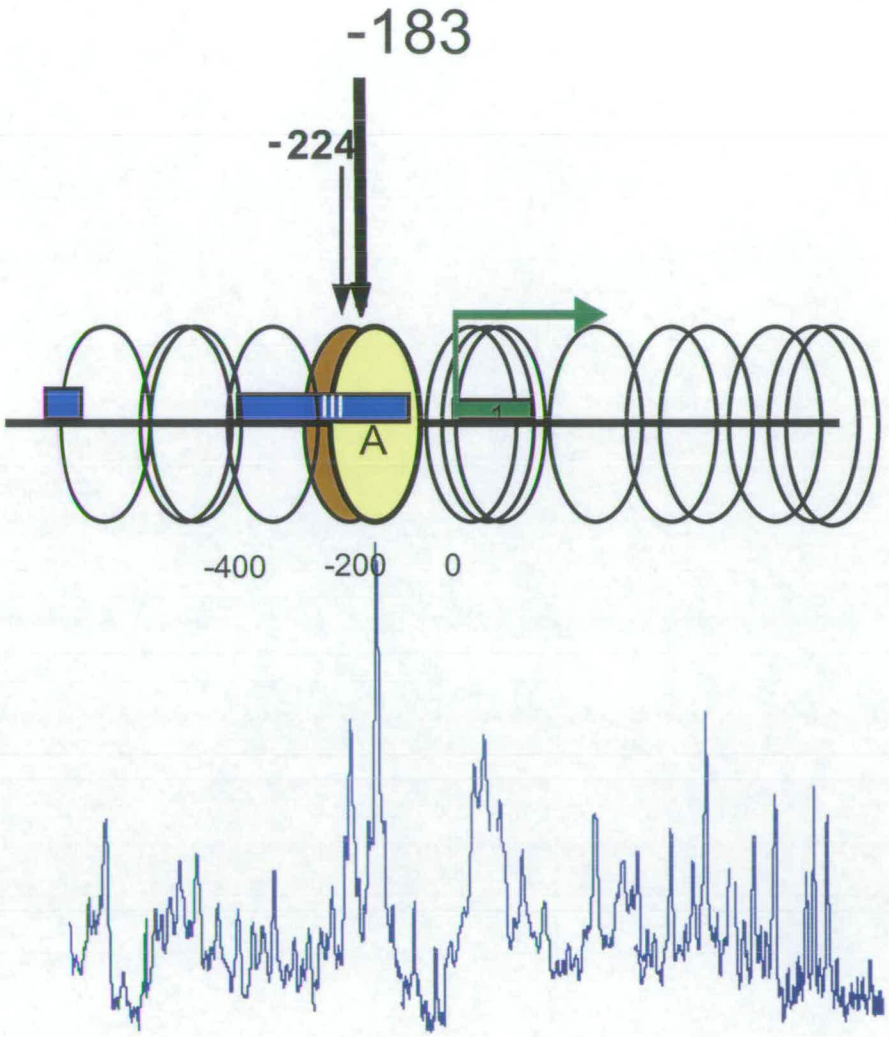


Figure 3.2. *In Vitro* Map of Nucleosome Positions over the BLG Gene.

Nucleosome positions derived from the *in vitro* analysis are represented as ovals over the BLG promoter and gene (top). The green arrow marks the transcription start site and the green box exon one. Blue boxes indicate the locations of DNaseI hypersensitive sites. A strongly positioned nucleosome, A highlighted in yellow, sits with its dyad at –183 bp. A second positioned nucleosome with a dyad at –224, nucleosome B, is highlighted in brown. Below the map is the signal intensity trace from the *in vitro* mapping experiment indicating the relative strength of each nucleosome position. This image was adapted from Figure 74 of (Boa, 1999).

bp from transcription start (Figure 3.3). This placement provides a separation of 185bp, which could accommodate a single nucleosome plus linker DNA. In fact this spacing corresponds closely to the genomic nucleosome repeat length found in mammary cells (Boa, 1999). The strong *in vitro* positioned site, nucleosome A, has a dyad estimated to be at –183bp, which would place it precisely between the –93 and the –278 Stat5 binding sites (yellow oval in Figure 3.3).

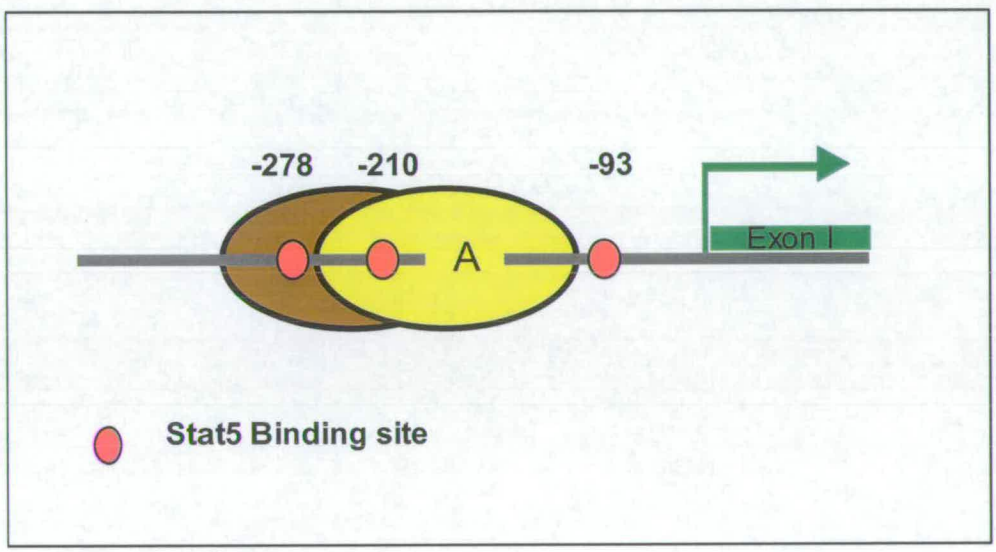


Figure 3.3. Detail of the *In Vitro* Positioned Nucleosomes on the BLG Gene.

Nucleosome position A is shown in yellow and the weaker position B in brown. Stat5 binding sites are represented by red circles. Nucleosome A is located between the Stat5 binding sites at –93 and –278 but covering the site at –210. Nucleosome B incorporates the –210 and the –278 Stat5 binding sites leaving only the –93 site accessible in linker DNA.

The *in vitro* map (Figure 3.2) also shows a second relatively strong nucleosome positioning site at –224, referred to hereafter as nucleosome B. This position overlaps nucleosome A and would lie over two of the Stat5 binding sites, leaving only the –93 site exposed in the linker DNA (brown oval in Figure 3.3). Of course only one of these alternative nucleosome positioning sites could be occupied in a cell at any one time.

The two alternative *in vitro* nucleosome positioning sequences nucleosome A and nucleosome B lie in the same region of DNA as, and may correspond to, the two alternative positions observed immediately before the transcription start site (marked by stick arrows in Figure 3.1) at the boundary between the regularly spaced array seen in the promoter of the active gene *in vivo*, and the disrupted nucleosome structure through the gene. These mapped nucleosomes also correspond to the region that is covered by HSIII, the presence of which correlates strongly with expression of BLG.

The *in vivo* nucleosome map shows that there is a distinctive chromatin structure over the proximal promoter of the BLG gene. It is possible that the *in vivo* mapped nucleosome pair immediately proximal to the transcription start site corresponds to the two *in vitro* mapped nucleosome positions A and B, suggesting a mechanism whereby the nucleosome positions could influence Stat5 binding, or alternatively a mechanism by which Stat5 could influence the nucleosome positions. Stat5 plays a major role in the regulation of BLG (Burdon et al., 1994b) and in the regulation of other milk protein genes. Two of the Stat5 binding sites are accessible, i.e. in linker DNA external to nucleosome A but in nucleosome B only site one is external.

3.2 Reconstitution of a Nucleosome onto the β–Lactoglobulin Promoter.

In this chapter, chromatin structure over the proximal region of the BLG promoter will be recreated in an *in vitro* reconstitution system, in order to study Stat5 binding to the promoter within a nucleosomal context. The reconstitution system employed is explained and the positions nucleosomes locate on the DNA fragment are examined.

3.2.1 Fragment Design and Rationale.

A fragment of the BLG promoter was chosen that incorporates the three Stat5 binding sites and the two prominent *in vitro* nucleosome positions (Figure 3.4). The fragment is denoted nAB for the two nucleosome binding sites.

nAB was produced by PCR using primers nAF and nBR. The product is 240bp long and runs from –312 to –73 with respect to the transcription start site (Figure 3.4). As each core binds 146bp of DNA only one core histone octamer should normally interact with a DNA molecule of this size. The fragment was designed as the minimum that would incorporate both of the alternative nucleosome positions, plus the three Stat5 binding sites, while remaining small enough to minimise the binding of multiple nucleosomes which can occur on fragments 260bp and larger (Tatchell and van Holde, 1978).

3.2.2 Reconstitution.

Core histones were reconstituted onto 5′ end-labelled nAB by redistribution from chicken erythrocyte donor chromatin using dialysis from high to low salt (section 2.8.4). At high concentrations of NaCl (2M), core histones are

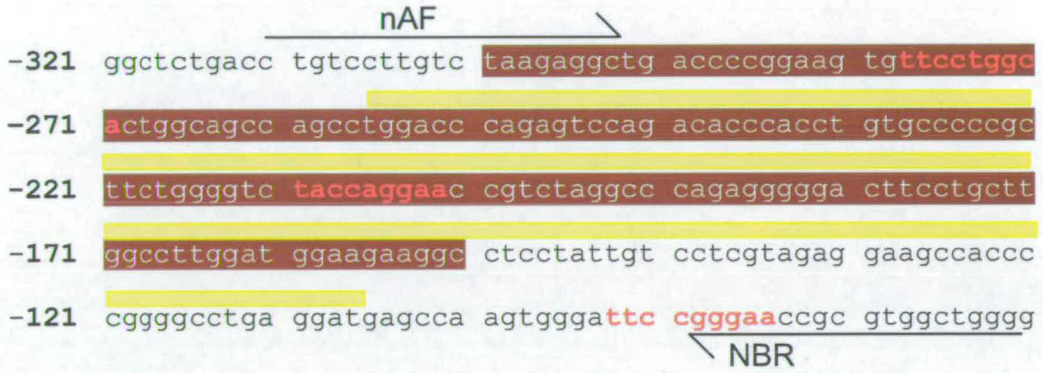


Figure 3.4 Sequence of the Probe nAB.

PCR primers nAF and nBR are indicated. The predicted nucleosome position A is marked by a yellow line above the sequence. Nucleosome position B is highlighted in brown. Stat5 binding sites are in red type.

dissociated from DNA. By slowly reducing the ionic strength from 2M through 0.8M to 10mM, histones reassociate and bind DNA (Tatchell and van Holde, 1977). During this process nucleosomes form preferentially on certain sequences (Beato and Eisfeld, 1997; Dong et al., 1990; Dong and van Holde, 1991; Yenidunya et al., 1994; Richard-Foy and Hager, 1987) although the precise parameters determining this process are not yet fully understood.

3.2.3 Positioning Isomers.

Reconstituted nucleosomes were analysed on 5% native polyacrylamide gels in TBE. Individual binding positions resolve on these gels according to their proximity to the centre of the DNA fragment. A centrally positioned nucleosome will be most retarded in its electrophoretic migration whereas a nucleosome positioned towards the end of the fragment will not be hindered to the same extent and will migrate more rapidly (Pennings, 1997). Thus individual nucleosome positioning isomers can be separated and isolated from a reconstitution. The strength of each nucleosome binding site is reflected by the intensity of the band observed.

3.2.4 Predicted Nucleosome Isomer Positions.

In a typical reconstitution onto nAB, three nucleosome positioning isomers are detected, labelled N1 through N3 (Figure 3.5). The slowest migrating band, N1, is also the most intense and therefore the most abundant. The slowest migrating nucleosome is predicted to be positioned most centrally on the DNA fragment. The strongest positioned nucleosome from the *in vitro* map, nucleosome A, is predicted to lie most centrally on nAB with DNA tails external to the nucleosome of 54 and 37bp (Figure 3.6C). Thus it is likely that N1 represents nucleosome A.

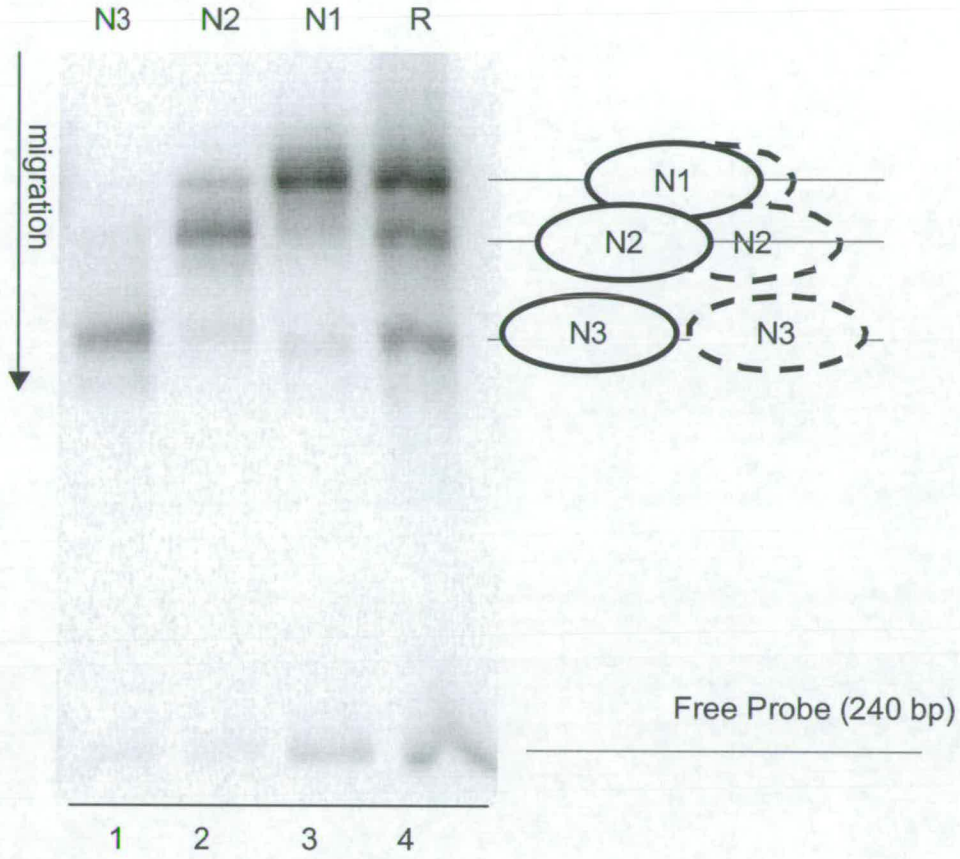


Figure 3.5 Nucleosome Positioning Isomers Resolved on a Non-Denaturing Polyacrylamide Gel.

Nucleosomes reconstituted onto nAB were separated on a 5% native polyacrylamide gel. The unfractionated reconstitute (R) is shown in lane 4, and purified isolated nucleosome positioning isomers N1-N3 in lanes 3-1. Schematic representations of the possible nucleosome positions from their migration in a native polyacrylamide gel are shown on the right hand side, these are based on the proximity to the centre of the DNA.

The nucleosome positioning isomer N2 migrates slightly faster than N1. N2 is probably positioned more towards a DNA end than N1 is. N3 migrates most rapidly through the gel suggesting this nucleosome positions most closely to the end of nAB. Nucleosomes exhibit a tendency to position at end positions; a reconstitution onto any DNA fragment will often show a tendency to produce an end positioned nucleosome. N3 may represent such an end position, or alternatively it may represent the second *in vitro* mapped nucleosome (B), which is predicted to have DNA tails external to the nucleosome of 13 and 78bp (Figure 3.6C). From their migration through a non-denaturing polyacrylamide gel either N2 or N3 could represent nucleosome B.

In this chapter I show that N1, and also N2, represent the *in vitro* mapped position A. A closer look at nucleosome A in Figure 3.2B shows that there may be two nucleosome positions flanking the main band by ~10 bp either side. N2 may be one or both of these shoulder positions. The third nucleosome positioning isomer, N3, is likely to be nucleosome position B whose positions predicts it lies close to the end of nAB. Nucleosome B, like nucleosome A contains flanking nucleosomes that show up in the *in vitro* map (Figure 3.2B).

3.2.5 Isolation of Nucleosome Positioning Isomers.

Individual nucleosome positioning isomers were isolated after separation on a native polyacrylamide gel by excision of gel slices containing the individual bands and elution into TEP_{10}. Samples of isolated positioning isomers run alongside the original reconstitution mixture show that in the low salt, low temperature conditions employed for elution the majority of nucleosomes remain at the same positions on the probe (Figure 3.5). Only a small fraction

of the probe became free of protein and redistributed between positioning isomers was minimal. A small percentage of both N1 and N2 does move to N3, and some N2 moves to N1. N3 is not observed to move. These "contaminating" bands may also reflect a failure to cleanly separate the positioning isomers, as N1 and N2 in particular migrate quite closely together in the gel. Nevertheless, the vast majority of the signal does remain in the correct bands.

3.3 Mapping of Nucleosome Core Particles.

In order to determine exactly where the Stat5 binding sites are in relation to each positioned nucleosome, the nucleosome binding positions on fragment nAB had to be mapped precisely. The positions of the two strong binding sites, N1 and N2, were mapped using two enzymatic approaches involving ExonucleaseIII (ExoIII) and micrococcal nuclease (MNase) with restriction digestion. The methodology behind each approach is distinct, thus the combined data from the two approaches should give a more reliable indication of the locations of nucleosome positions on nAB.

3.4 Exonuclease III Mapping.

ExonucleaseIII (ExoIII) removes nucleotides from the 3'OH end of double stranded DNA in a stepwise manner (Richardson et al., 1964; Richardson and Kornberg, 1964). It displays a slight sequence specificity of C>A=T>G (Linxweiler and Horz, 1982), thus a partial digest of naked DNA shows a unique transient pattern of pause sites. The presence of a positioned nucleosome on a DNA fragment causes ExoIII to exhibit a notable pause at the nucleosome-DNA boundary. Additional digestion within a nucleosome does occur with a distinct 10bp pattern (Riley and Weintraub, 1978; Kefalas et al., 1988).

In order to harness ExoIII to map nucleosome positions on DNA, fragments were produced that were labelled uniquely at one 5′ end in order to map ExoIII pause sites to each fragment end separately. Unique cutting restriction enzyme sites were incorporated into the ends of nAB in order to be able to selectively remove one labelled end. A PCR primer was designed with a *Pvu*II site in the sequence of the upstream primer nAF to give the primer nApF (Figure 3.6B). Similarly an *Rsa*I site was incorporated into the downstream PCR primer nBR to give primer nBrR (Figure 3.6B). Digestion of the resulting PCR product with either *Rsa*I or *Pvu*II after 5′ end labelling results in DNA uniquely labelled at only one end (Figure 3.6A (*Rsa*I) and (*Pvu*II)). These enzymes were chosen because they leave DNA with no overhangs after digestion; ExonucleaseIII cannot digest single stranded DNA, so restriction enzymes that would leave a 3′ overhang were avoided as these are very poor substrates for the enzyme. These fragments are subtly different from the fragment nAB (Figure 3.6A and Table 3.1) used for the original reconstitution, and as such may be expected to behave differently in a reconstitution. The ratio of DNA tails that are external to the nucleosome indicates how centrally on a given fragment a nucleosome sits; a nucleosome with a tail ratio of 1.0 is positioned right in the middle of a fragment. The higher the ratio, the closer to the end of the fragment a nucleosome positions.

Digestion of 5′ end labelled nApBr with *Pvu*II results in a fragment, called (*Pvu*II), labelled only once, at its downstream end. This can be used to map the upstream end of a positioned nucleosome by ExoIII digestion. In contrast, *Rsa*I digestion results in DNA labelled only at the upstream end and this fragment, (*Rsa*I), is used to map downstream boundaries of nucleosomes.

Figure 3.6. Comparison of Nucleosome Positions on nAB with nApBr.

A schematic representation of fragments nAB and nApBr and their relative ends (**A**). Positions of primers are shown relative to nucleosome A and nucleosome B. Full length PCR products nAB and nApBr, as well as the products of restriction digest of nApBr are shown as labelled. More information on each fragment mentioned can be found in Table 3.1. Primer sequences used are displayed in **B**. The downstream primer sequence shown has been converted to how it would appear on the upper strand, DNA will actually be labelled on the lower strand. Restriction sites incorporated into the primers are underlined, and cleavage points indicated by stars (*). The lengths of DNA external to each nucleosome, A or B, are shown in **C.** Nucleosome A on fragment nAB has DNA tails of 54 and 37 bp; the ratio of these is 1.46. The ratio of the lengths of the DNA tails indicates how centrally a nucleosome is positioned on the fragment. This ratio was affected by the changes introduced into the fragment by use of the primers that introduce restriction enzyme sites. The lengths of tail DNA external to the nucleosome core is shown for each fragment, and the ratio of these is shown in the column on the right hand side.

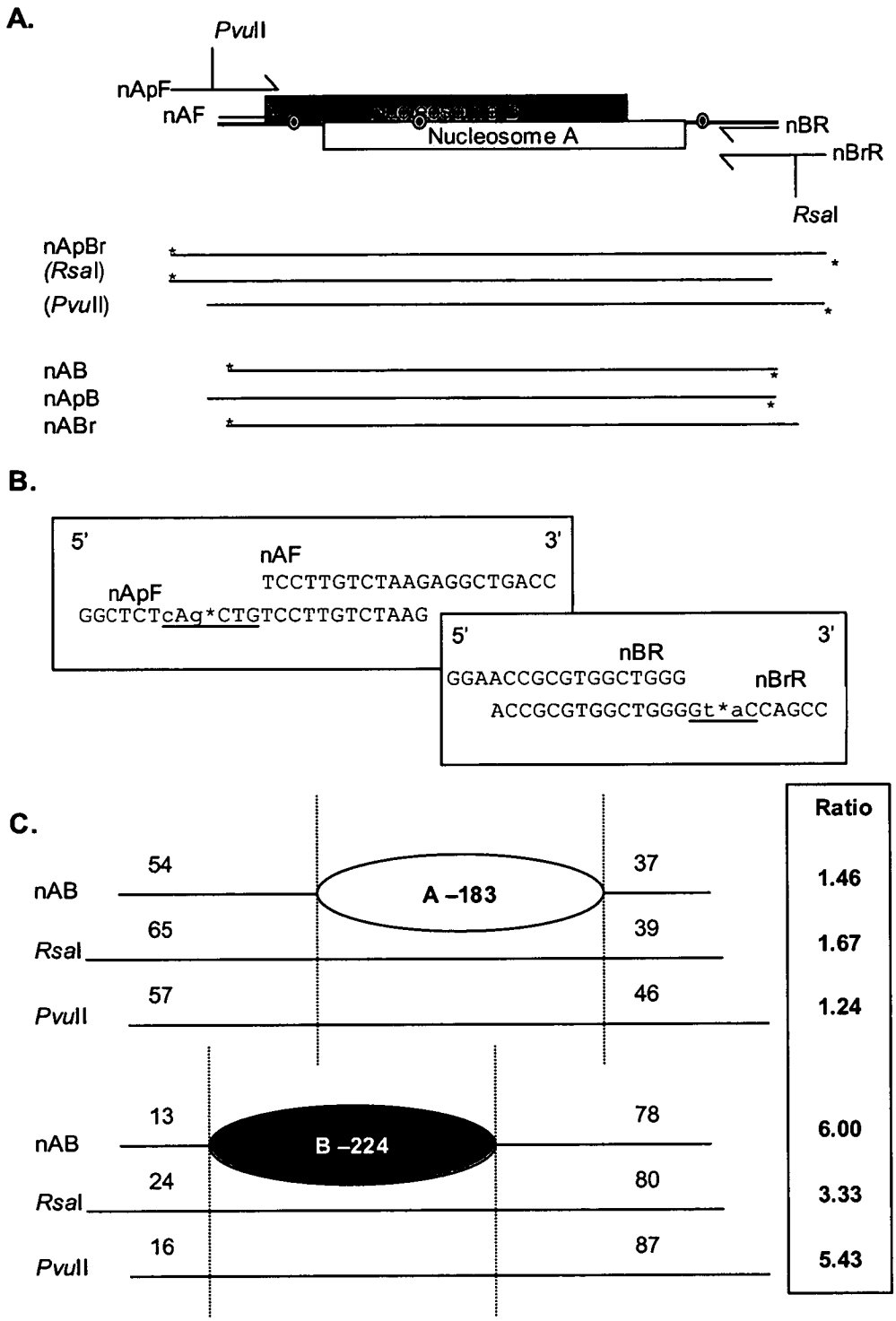


Figure 3.6. Comparison of Nucleosome Positions on nAB with nApBr.

Table 3.1 Description of Fragments used for Reconstitution.

Fragment name	Length (bp)	Start	End	Primers	Digest
nAB	240	-312	-73	nAF : nBR	-
nApBr	261	-324	-64	nApF : nBrR	-
nApB	252	-324	-73	nApF : nBR	-
nABr	249	-312	-64	nAF : nBrR	-
(*Rsa*I)	254	-324	-71	nApF : nBrR	*Rsa*I
(*Pvu*II)	252	-315	-64	nApF : nBrR	*Pvu*II
nApB	243	-315	-73	nApF : nBR	*Pvu*II
nABr	242	-312	-71	nAF : nBrR	*Rsa*I

A second set of digests was carried out to demonstrate complete cleavage of labelled ends. The size of the cleaved end fragment is too small in relation to the whole fragment to allow verification of its removal directly (Figure 3.7, compare lanes 4, 5 and 6). Removal of the ends was verified by double digestion with internally cutting enzymes *Acc*I (Figure 3.7 lanes 1 –3) and *Sma*I (Figure 3.7 lanes 7 – 9). *Acc*I cleaves the full-length sequence into two fragments of 100 and 142bp (lane 3 and 3.7B), but in a double digest with either of *Pvu*II (lane 2) or *Rsa*I (lane 1) only one fragment is labelled and shows up on the gel. *Sma*I cleaves the uncut sequence into 190 and 52bp (lane 9 and 3.7B). With a double digest only one product is seen indicating the ends have successfully been removed (Figure 3.7 lanes 7 and 8). Figure 3.7 B shows the positions of the restriction digest sites, and the calculated fragment sizes.

3.4.1 Characterising Reconstitutes to be used in ExoIII Mapping.

Nucleosomes were reconstituted onto the uniquely end-labelled fragments described above and individual positioning isomers were isolated (Figure 3.8). The pattern of nucleosome migration on (*Rsa*I) and (*Pvu*II) show some subtle differences from each other (Figure 3.8). Reconstitutes prepared on (*Pvu*II) have a similar pattern to that seen with reconstitutes prepared on the normal nAB probe, but two nucleosome positions are observed in respect of the fastest migrating band N3 (compare nAB in Figure 3.5 lane 4 to (*Pvu*II) in Figure 3.8 lane 4). The reconstitute on (*Rsa*I) also has two bands corresponding to N3, plus an extra band migrating closer to the central N2 position (Figure 3.8 lane 5). In (*Rsa*I), the isomer at the N1 position also appears to migrate slightly faster than it does in (*Pvu*II). An attempt to isolate all four positions from the reconstitution onto (*Rsa*I) was unsuccessful and resulted in nucleosome N1 isomers being split between two fractions.

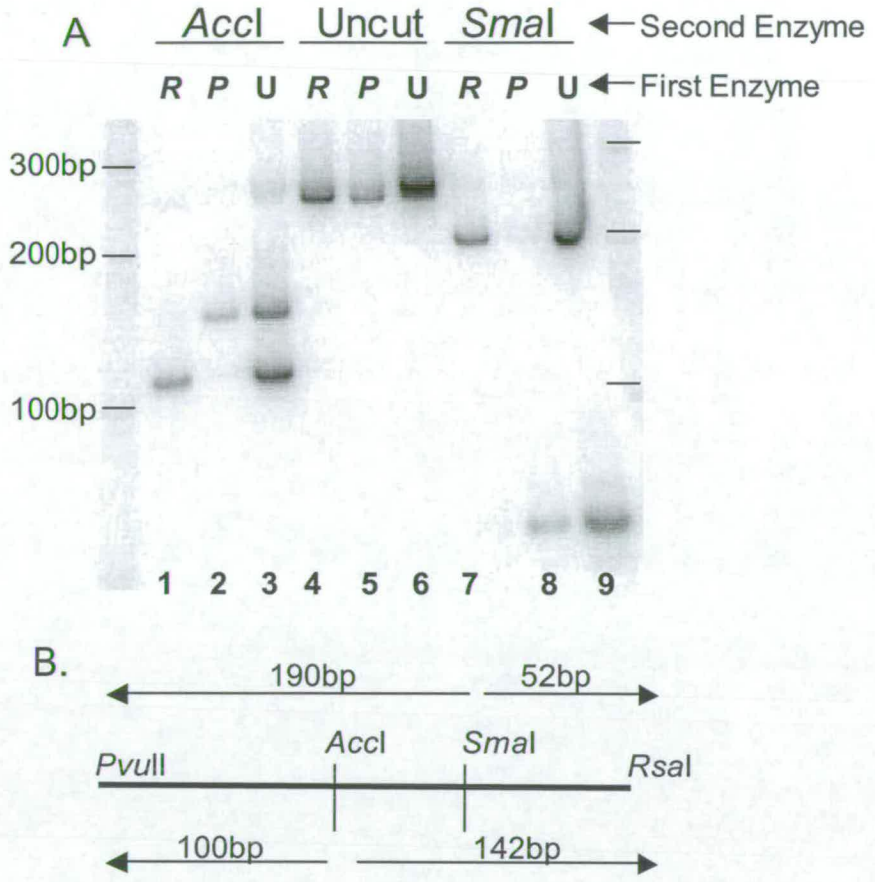


Figure 3.7. Restriction Enzyme Digests to Verify Correct Removal of the Upstream or Downstream Labelled Ends of nApBr.

A. Products from restriction digests of 5' end-labelled nApBr were separated on an 8% polyacrylamide gel. Labelled ends were removed by digestion either with *Rsa*I (R) (lanes 1, 4 and 7) or *Pvu*II (P) (Lanes 2, 5 and 8) DNA without either end removed (U) is also shown (lanes 3, 6 and 9). Subsequent digestion with a second enzyme that cleaves internally in the fragment verifies the removal of each end. A schematic diagram showing the cleavage sites for each enzyme is shown in **B**.

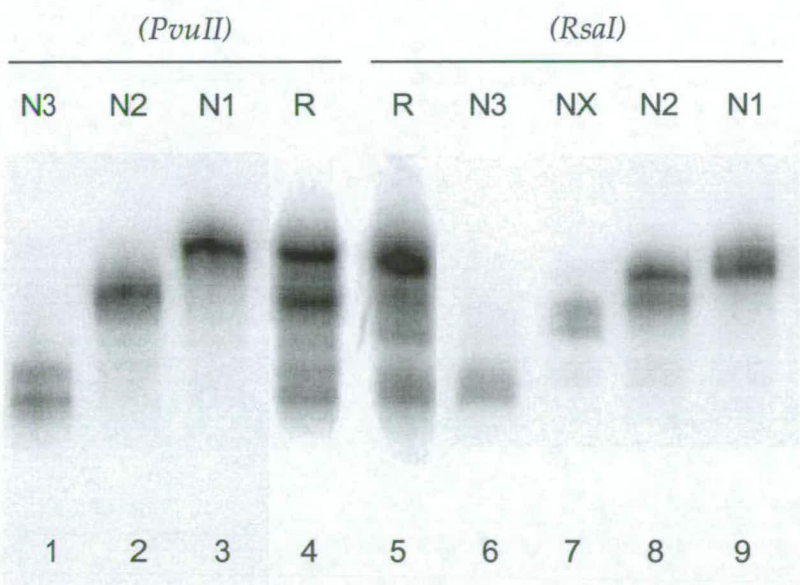


Figure 3.8 Reconstitution of Nucleosomes onto (*Rsa*I) and (*Pvu*II).

Lane 4 is an unfractionated reconstitute (R) on (*Pvu*II), and lanes 1-3 are the individual positioning isomers isolated from this material. Lane 5 is an unfractionated reconstitute on (*Rsa*I), and lanes 6-9 are the individual positioning isomers isolated from this material.

The N1 fraction from the (*Rsa*I) reconstitute contains only the slowest migrating nucleosome N1, but the N2 fraction contains bands corresponding to isomers N1 and N2. In the subsequent ExoIII digests (Figure 3.12) the N2 fraction will contain digest products of both the N1 isomer and the N2 isomer, so only bands seen in the (*Rsa*I) N2 fraction that are not present in the (*Rsa*I) N1 fraction must belong to position (*Rsa*I) N2. No attempt was made to separate the two fastest migrating nucleosomes (N3) in either of the nApBr reconstitutions.

The differences in migration between the (*Pvu*II) and the (*Rsa*I) reconstitutes may be explained by the extra nucleotides present at the ends of the sequence resulting from the use of primers incorporating the *Pvu*II and *Rsa*I restriction sites (Figure 3.6). Using the previously mapped positions of nucleosomes A and B (Figure 3.2) as an example, nucleosome A, predicted to sit at –183, would have free DNA tails of 54 and 37bp on a reconstitute on nAB. The ratio of the free DNA tails gives an indication of how centrally on the fragment a nucleosome lies. The closer this ratio is to 1 the more centrally the nucleosome lies. The ratio of the free DNA ends for nucleosome A on fragment nAB is 1.46 (Figure 3.6C), indicating that it would be a fairly central position, and thus is likely to migrate relatively slowly through a polyacrylamide gel. The same nucleosome A reconstituted on (*Pvu*II) would be positioned even closer to the centre of the fragment, having a DNA end ratio of 1.24. The reverse is true with a reconstitute on (*Rsa*I) as the ratio rises to 1.67 (Figure 3.6C). Nucleosome A on (*Rsa*I) would migrate slightly faster than on nAB, and on (*Pvu*II) it would migrate slightly slower than on nAB. Thus the predicted migration rates for Nucleosome A on the (*Rsa*I) and the

(*Pvu*II) fragments, shown in Figure 3.6, are consistent with the difference seen in migrations for the N1 isomer in Figure 3.8.

The main N2 isomer band appears to migrate at the same level in reconstitutes on all three fragments. However in reconstitutes on (*Rsa*I) the band reflecting the N2 isomer is of lower intensity than the N2 isomer bands in reconstitutes on either (*Pvu*II) or nAB. A faster migrating band of similar intensity to the N2 isomer is also present in (*Rsa*I), designated NX (Figure 3.8). The total amount of signal in the (*Rsa*I) N2 and NX nucleosome isomer bands approximately equals the amount of signal contained in the (*Pvu*II) and nAB N2 isomer bands, suggesting NX may be a component of N2. This could indicate that the subtle differences in the sequence at the end of the fragment result in a proportion of the nucleosomes in N2 positioning at a different site. However these changes in the DNA sequence are at the very ends of the fragment and as such are predicted only to interact with an end-positioned nucleosome. The migration of N2 through a polyacrylamide gel does not imply it is a nucleosome positioned at the end of a fragment, making it unlikely that interactions with the sequences at the end of the fragment could result in a completely new nucleosome position. Alternatively it could indicate the possibility that N2 actually harbours two separate nucleosome positions, which are not resolved in reconstitutes on nAB or (*Pvu*II) but the subtle difference in free DNA tail length in (*Rsa*I) allows the two isomers to resolve.

N3, predicted to be nucleosome B, migrates as two bands in reconstitutes on both the (*Rsa*I) and (*Pvu*II) fragments but not in reconstitutes on nAB. The ratio of free DNA lengths for nucleosome B on (*Pvu*II) and nAB are similar at

6.00 and 5.43 (Figure 3.6). However this is much reduced in (*Rsa*I) with a ratio of 3.33. This difference may be able to account for the isomer NX migrating between N2 and N3 in (*Rsa*I). However the fact that the two bands associated with N3 are present in reconstitutes on both (*Pvu*II) and (*Rsa*I) argues against this redistribution (Figure 3.8).

It is unclear how these subtle differences in fragment length might affect the migration of N3. It is possible that differences are not resolved for nucleosomes positioned towards the end of a fragment. It is likely that N3 is composed of multiple positioned nucleosomes, some of which are directed by the DNA sequence and some by end effects. End positioned nucleosomes can lie at either end of the DNA fragment, this may explain the similar positions observed for both (*Pvu*II) and (*Rsa*I).

The subtle differences in migration of the N1 isomer have been explained by changes in free DNA lengths. Analysis would be much more straightforward if nucleosome isomer patterns were identical for each fragment. Different combinations of primer pairs were employed to minimise the difference produced at the ends of the fragments. Primer nAF was paired with primer nBRr to produce the fragment nABr, which could be cleaved by *Rsa*I. Similarly primer nAFp was paired with primer nBR to produce nApB, which could be cleaved by *Pvu*II (Figure 3.6 and Table 3.1). These combinations only differ by two or three base pairs after digestion with the relevant enzyme.

Reconstitution experiments on nApB and nABr (Figure 3.9) look essentially the same as a reconstitution on nAB (Figure 3.5) with the exception that the

nApB N1 isomer migrates as two bands. This is unlikely to represent a different nucleosome position caused by changes in the DNA sequence, as these changes are not within the region that is predicted to be covered by the nucleosome N1. The mutations introduced to add the *Pvu*II restriction site, even after digestion, result in addition of an extra three base pairs (CTG) to the upstream end of the fragment compared to the sequence of nAB. This may be enough to resolve different nucleosome positioning isomers within N1, although why these were not resolved in reconstitutes on (*Pvu*II) and (*Rsa*I) is unclear. Verification of the removal of the labelled end was not carried out on nApB and nABr, so the band could be explained by incomplete digestion. Notably both nABr and nApB have only one nucleosome isomer at N2 indicating that the differences seen at this position in (*Pvu*II) and (*Rsa*I) reconstitutions were due to the extra lengths of DNA. ExonucleaseIII experiments described below are on nucleosomes isolated from reconstitutions onto (*Pvu*II) and (*Rsa*I).

3.4.2 Positioning Stability Under ExoIII Conditions.

To show that nucleosome positions are not affected by the conditions used in the ExoIII digests, isolated positioning isomers were treated with ExoIII and their separation on native polyacrylamide gels was compared to untreated nucleosomes. An unfractionated reconstitute treated under ExoIII conditions, but in the absence of ExoIII, looked identical to an untreated reconstitute (Figure 3.10 compare lanes 1 and 2). There was also no nucleosome movement detected under these conditions among isolated position isomers (data not shown). This demonstrated that the buffer and temperature change has no noticeable affect on nucleosome positioning.

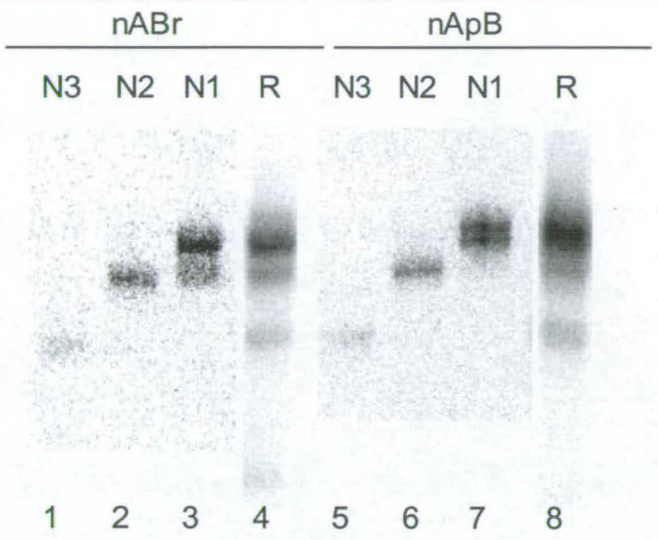


Figure 3.9 Reconstitution onto nApB and nABr, cut Respectively with *PvuII* or *RsaI*.

Lane 4 is an unfractionated reconstitute (R) on nABr, and lanes 1-3 are the individual isolated positioning isomers from this material. Lane 8 is an unfractionated reconstitute on nApB and lanes 5-7 are the individual isolated positioning isomers from this material.

Isolated positioning isomers were also treated with either 200 or 800 units/ml ExoIII and again separated on a polyacrylamide gel in TBE. Nucleosome positioning isomers treated with the lower concentration of ExoIII migrate essentially the same as the untreated samples (Figure 3.10 compare lanes 3 with 4, 6 with 7, and 9 with 10). Only the N2 nucleosome isomer displays redistribution, to a position closely resembling the NX nucleosome isomer observed in Figure 3.8. Digestion with ExoIII at the higher concentration of 800 u/ml causes nucleosomes to migrate in a different pattern (Figure 3.10 compare lane 3 with 5, lane 6 with 8 and lane 9 with 11). As shown by the extent of digestion in Figures 3.11 and 3.12, at this concentration ExoIII will have progressed well into the nucleosome, probably up to the dyad axis. This means the nucleosome structure is unlikely to be as stable as that of an intact nucleosome having only one turn of double stranded DNA to hold it together. On the whole nucleosomes appear to maintain their position, an observation supported by the presence of a 10bp cutting pattern observed when the digest products are separated on a denaturing polyacrylamide gel (Figure 3.11 and 3.12). This strongly suggests a static, translationally positioned nucleosome.

3.4.3 ExoIII Mapping of Upstream Nucleosome Boundaries.

In order to map upstream nucleosome boundaries, isolated nucleosome isomer populations recovered from a reconstitute on fragment (*Pvu*II) were digested with increasing amounts of ExoIII. Samples of unreconstituted (naked) DNA were also digested, but with lower amounts of ExoIII. Reactions were stopped after 15 minutes at 20°C and DNA products were separated on a 10% denaturing polyacrylamide gel. The length of each labelled DNA fragment corresponds to the distance in base pairs of each

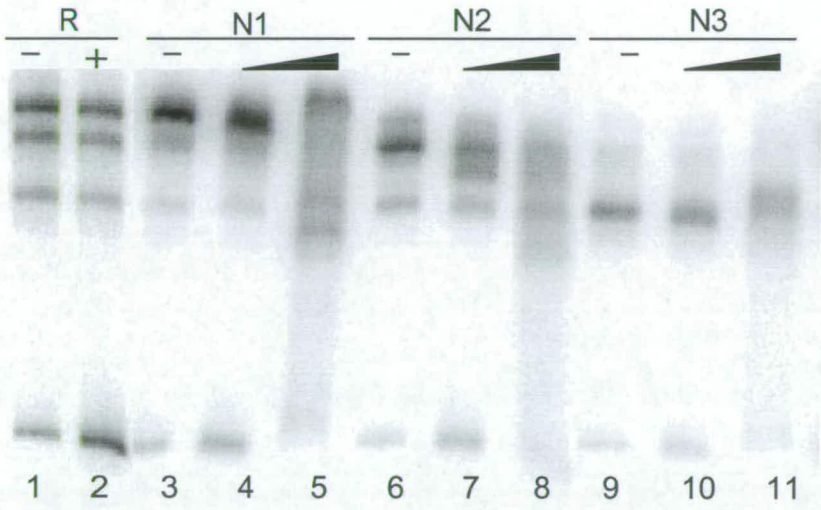


Figure 3.10 Nucleosome Positioning Stability under ExoIII Conditions.

Nucleosome positions on uniformly labelled WT nAB were analysed under Exonuclease III conditions to demonstrate nucleosome positioning isomers are not affected by treatment with ExoIII. Lane 1 shows an unfractionated reconstitute in TEP_{10}, maintained at +4°C. Under these conditions nucleosome positions remain stable. Lane 2 is the same reconstitute incubated in ExoIII buffer in TEP_{10} at 20°C for 15 minutes. Lanes 3, 6 and 9 are untreated nucleosome positions N1, N2 and N3 respectively in TEP_{10}. Lanes 4, 7 and 10 have been digested with 200 units/ml and lanes 5, 8 and 11 with 800 units/ml ExoIII.

ExoIII pause site from the labelled end, which is located at –64bp from the transcription start site (Figure 3.13).

The ExoIII map of nucleosome positions derived from N1, N2 and N3 from (*Pvu*II) is shown in Figure 3.11. Analysis of the band sizes was carried out by reference to 10bp markers (Invitrogen).

The ExoIII digest of N1 shows bands of 238, 227 and 220 nucleotides (nt), but these are also clearly present in the naked DNA digests and thus are likely to be DNA sequence specific pause sites. The next major N1 specific pause site produces a fragment of 208 nt. This band marks the beginning of a group of pause sites, spaced with a 10nt periodicity, at 208, 198, 188nt etc. This pattern is characteristic of ExoIII cutting of DNA associated with a nucleosome (Riley and Weintraub, 1978; Kefalas et al., 1988), indicating that the 208nt band marks the boundary of the N1 position. The upstream boundary of the N1 position is therefore 208nt from the *Rsa*I end of the fragment; i.e. at –272 bp from the transcription start site (Figure 3.13). The N1 upstream nucleosome boundary and pause sites, with respect to the transcription start site, are marked in red in Figure 3.11. The fourth band in the 10nt pattern appears to reflect a slightly stronger pause site, this may reflect a structural feature of this nucleosome.

The ExoIII digest pattern of N2 is not as straightforward as that of N1 (Figure 3.11). The same DNA sequence specific pause sites are seen as in N1, plus there are pause sites that give rise to prominent fragments of 200 and 190bp. These sites are present at only very low ExoIII concentrations. The same pause sites are seen in the naked DNA digest but here they are not nearly as

Figure 3.11 (*Pvu*II) ExonucleaseIII Protection Mapping.

Isolated nucleosome positioning isomers N1, N2 and N3 on (*Pvu*II) were digested with 200, 800 and 3000 u/ml ExoIII and the corresponding control DNA digested with 0, 50, 100 and 400 u/ml ExoIII. Products were separated on 10% denaturing gels. Marker sizes in nt are marked in black. Features of the N1 nucleosome are marked in dark red; numbers refer to positions on the gene in base pairs from the transcription start site. An arrow on the left side marks the boundary of N1 at -272, and dashed lines on the left show the 10bp periodicity of digestion that is characteristic of a positioned nucleosome. Features of the N2 nucleosome are marked in dark blue on the right hand side of the gel. Again numbers refer to base pairs from the transcription start site. The arrow on the right marks the boundary of N2 at –242. The thick black line marks a strong nucleosome specific cut site, not associated with the N2 nucleosome boundary. Dashed lines on the right mark the internal 10bp pattern. There is a potential N3 specific pause site at 228nt which corresponds to a position on the gene of –292, this is marked by a grey arrow. m is a 10bp ladder. Figure 3.11 is on the following page.

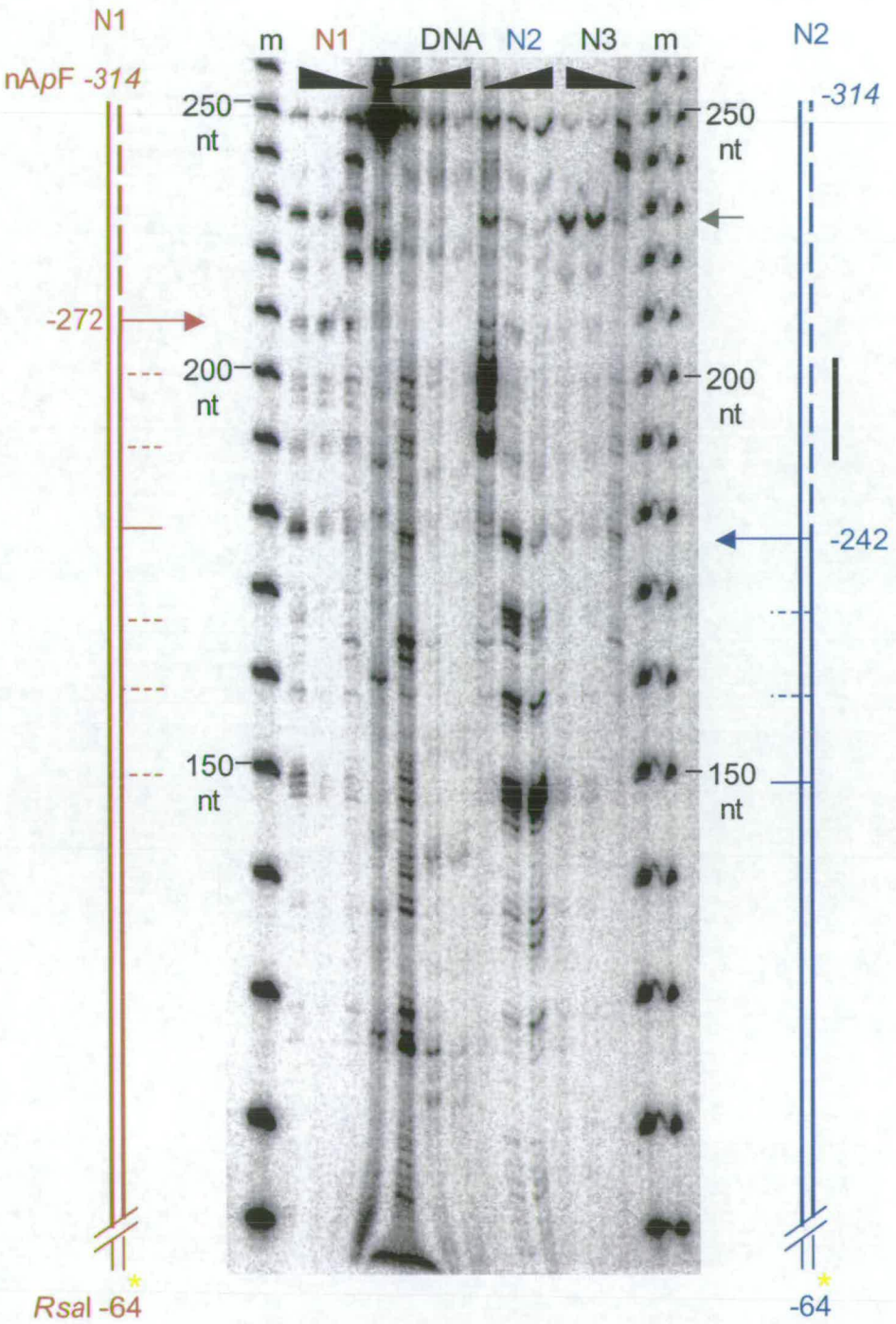


Figure 3.11 (*Pvu*II) ExonucleaseIII Protection Mapping.

strong. These DNA-specific ExoIII pause sites may be strengthened by interactions of DNA external to the nucleosome core with the nucleosome surface. The first non sequence-specific pause site present at the higher enzyme concentrations produces a band of 177bp. This site again marks the beginning of a 10bp pattern, which extends for 30bp into the nucleosome to a second strong pause site in a similar pattern to the N1 digest. The upstream boundary of N2 is therefore 177nt from the *Rsa*I end of the fragment i.e. at –242nt from start (Figure 3.13). The upstream nucleosome boundary and pause sites, referenced to the transcription start in N2 are marked in blue in Figure 3.11.

A digest of N3 is also included in this gel, but no non-sequence-specific bands were observed in this analysis. The amount of signal in N3 is much less than that in N1 or N2, nevertheless an ExoIII pause site at 228nt is detected (marked by a grey arrow in Figure 3.11). This band is also detected in naked DNA digests and in N1 and N2 digests, but in N3 is much stronger when compared to the total signal present. It is possible that this represents a N3 nucleosome boundary that happens to coincide with a DNA specific pause site thus placing the upstream boundary of N3 at –292bp from the transcription start site.

3.4.4 ExoIII Mapping of Downstream Nucleosome Boundaries.

The downstream ends of the nucleosome positions N1 and N2 were mapped on the upper strand from a reconstitution onto (*Rsa*I) (Figure 3.12). Digests of NX and N4 were also carried out but no pause sites were detected (data not shown).

Figure 3.12 *Rsa*I ExonucleaseIII Protection Mapping.

Isolated nucleosome positioning isomers N1 and N3 were digested with 200, 800 and 3000 u/ml ExoIII and the corresponding control DNA digested with 50, 100 and 400 u/ml ExoIII. The lane beside the marker at the right hand side is undigested DNA. Products were separated on 10% denaturing polyacrylamide gels. Marker sizes in nt are marked in black. Features of the N1 nucleosome are marked in dark red; numbers refer to position on the gene in base pairs from the transcription start site. An arrow on the left side marks the boundary of N1 at –123, and dashed lines on the left show the 10bp periodic digestion characteristic of a positioned nucleosome. Features of the N2 nucleosome are marked in dark blue on the right hand side of the gel. Again numbers refer to base pairs from the transcription start site. The dashed line on the right hand side marks the predicted boundary of N2 at –93. m is 10bp ladder. Figure 3.12 is on the following page.

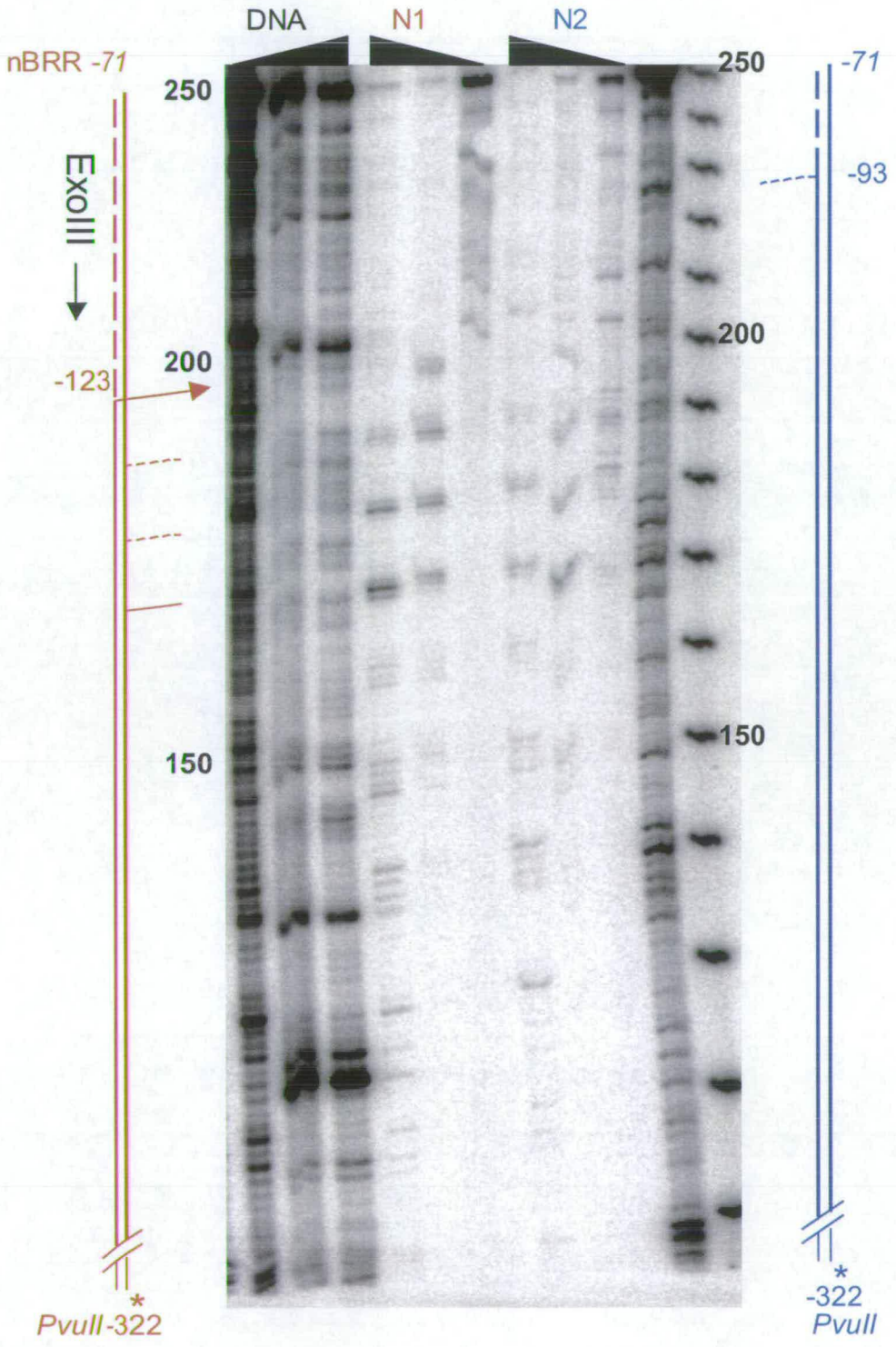


Figure 3.12 (*Rsa*I) ExonucleaseIII Protection Mapping.

The ExoIII digest of N1 has strong nucleosome specific bands of 199, 189, 179and 169nt. Again these follow a 10bp pair periodicity from a strong pause site at 199nt; the boundary of nucleosome position N1 is therefore 199nt from the upstream labelled end (–322) mapping the downstream boundary of N1 to –123bp from start (Figure 3.13).

The ExoIII digest of N2 displays a very similar pattern to that of N1. A close look at the separation of the isolated positioning isomers from the reconstitute on (*Rsa*I) (Figure 3.8) shows that the major band in both the N1 and N2 isolates migrates at the same rate and is probably the same band. There is a weaker band in the N2 analysis that migrates at the level predicted for, and probably is, the actual N2 position. Assuming a nucleosome protects the accepted 146bp of DNA, the predicted downstream boundary would be at –93 and should show up in Figure 3.12 as a band of 229nt. This point is marked in Figure 3.12 beside the N2 digest. There is a sequence specific pause site at this position which may conceal an N2 specific band. The signal in the N2 band is weak compared to the N1 band in the (*Rsa*I) reconstitute (Figure 3.8 compare lanes 8 and 9).

3.4.5 Exonuclease III Mapped Nucleosome Positions.

The N1 nucleosome position isomer has been mapped at both the upstream and the downstream boundaries, resulting in a protected DNA length of 149bp. Previous ExoIII digests have shown between 143 and 147bp DNA protected by a nucleosome (Linxweiler and Horz, 1985; Kefalas et al., 1988). Although the figure of 149bp is slightly larger, it is still in agreement with these values. Interactions between non-nucleosomal DNA and the histone core, could account for the additional protected DNA. There is a run of 4 G bases from –120 to –117 nucleotides from the transcription start site. This is 4

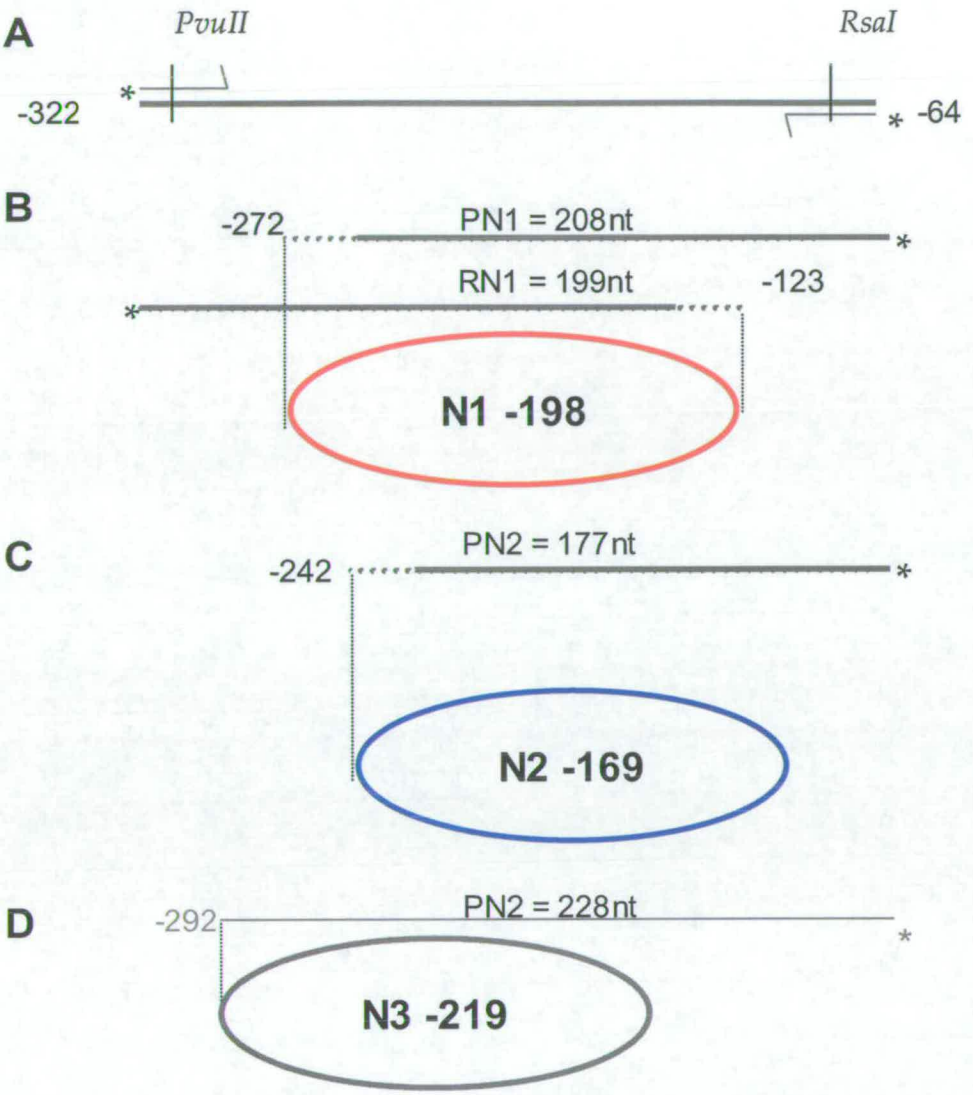


Figure 3.13 Nucleosome Positions Mapped by ExonucleaseIII Digestion.

Data from the ExoIII upstream and downstream mapping experiments are summarised. The fragment used is detailed in **A** with the ends marked with respect to the transcription start site of BLG. The fragment length detected in Figures 3.11 and 3.12 is marked in nt. The two ends of N1 (in red) have been mapped resulting in a nucleosome position covering 149bp of DNA (**B**). Position N2 (in blue) has only been mapped at one end. However the dyad can still be calculated from this, assuming a nucleosome covers 146bp of DNA. This positions the dyad at –169bp (**C**). N3 has been mapped at the upstream boundary, resulting in a nucleosome positioned at –219.

nt external to the downstream boundary of N1 and may be sufficient to cause the ExoIII to pause outside the core. A strong DNA sequence specific pause site corresponding to this run can be detected in Figure 3.12 migrating just above the 200nt marker. Assuming that each end is digested equally towards the nucleosome boundary, the centre or dyad of the nucleosome position N1 lies at –198bp from transcription start (Figure 3.13).

The nucleosome positions calculated by ExoIII digestion are summarised in Figure 3.13. Position N1 has been mapped at both boundaries and lies with its dyad at –198bp from the transcription start site. Nucleosome position N2 has been mapped at only the upstream boundary. If one assumes that the core DNA is 146 bp, the mapped upstream end would position the dyad at –169bp. By the same argument, N3 would position at –219bp.

The upstream boundary of N2 displays a tendency indicative of non core DNA associating with the nucleosome, as reflected by weak pause sites that are not strong enough to represent a nucleosome boundary (Figure 3.11). These interactions cover 20-30bp of DNA upstream of the nucleosome. This DNA separates the N2 nucleosome at –169bp from the N1 nucleosome at –198bp and may represent a link between the N1 and the N2 positions. A corresponding sequence with weak affinity for a nucleosome could exist at the downstream end of N2, explaining the lack of detection of a clear boundary here.

Exonuclease III mapping has produced positions for N1, N2 and N3. Mapping with a second technique should be able to reinforce these results. Currently either or both of N1 and N2 could be interpreted as the strong

positioned nucleosome A, mapped by monomer extension (Boa, 1999; Gencheva and Allan, 2005), as its position at –183 lies almost exactly between –169 and –198.

3.5 Micrococcal Nuclease and Restriction Digest Mapping.

Micrococcal nuclease (MNase) is a useful tool for the study of chromatin structure, as well as in the preparation of bulk chromatin. Reconstituted nucleosome positions were mapped by the method previously described (Meersseman et al., 1991; Yenidunya et al., 1994).

In this approach core DNA is prepared from a reconstitute by MNase digestion. DNA protected by a nucleosome is not readily digested by MNase (Noll and Kornberg, 1977; Horz and Altenburger, 1981; Alexander et al., 1961). This core DNA is digested with unique-cutting restriction enzymes, the unique pair of bands produced for each core particle isomer allows calculation of the position of the nucleosome boundaries with respect to the restriction site. Two unique restriction sites are required to absolutely map each core DNA with respect to its position on the gene.

3.5.1 Production of Core Particle DNA.

Fragment nAB was produced by PCR using primers nAF and nBR (Table 3.1) in a reaction including α^{32}P dATP, resulting in a probe labelled uniformly throughout its length.

Optimum conditions for digestion to core particle DNA were obtained for each reconstitution reaction by testing a small fraction of each at various time points. A typical MNase time course is shown in Figure 3.14. In this instance

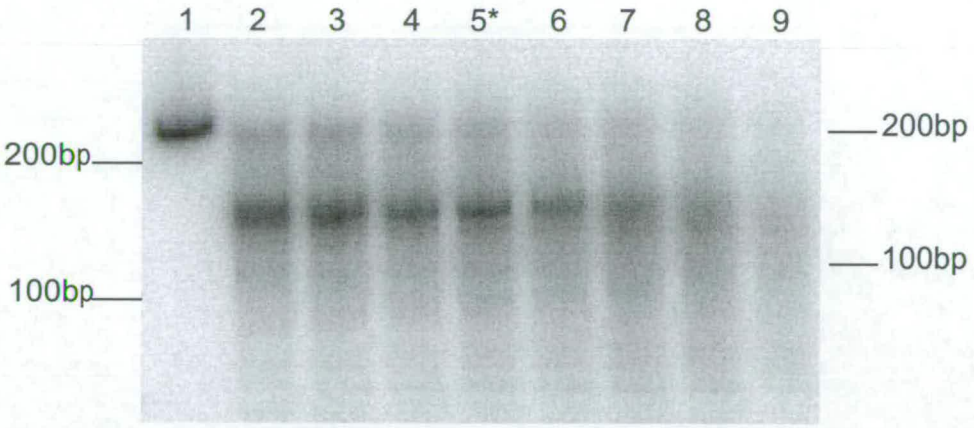


Figure 3.14 MNase Time Course.

Reconstituted DNA samples digested on ice for 20 minutes then at 21°C for 0 (lane 2), 0.5 (lane 3), 1 (lane 4), 2 (lane 5), 3 (lane 6), 4 (lane 7), 5 (lane 8), and 6 minutes (lane 9). Undigested DNA is run in lane 1.

DNA was first digested on ice for 20 minutes. This was followed by digestion at room temperature (21°C) for 0, 0.5, 1, 2, 3, 4, 5 and 6 minutes. At this latter temperature MNase is able to trim to the core particle boundary by virtue of its exonuclease activity. The time point of 2 minutes digestion at room temperature was chosen for the large-scale core particle preparation (Figure 3.14, lane 5). Digestion for longer periods of time resulted in MNase beginning to invade into the nucleosome and loss of the protected core DNA.

A mixture of nucleosome positions from a whole reconstitution reaction, rather than isolated individual positions was used in this experiment to maximise the available signal from weak nucleosome positions. The core DNA produced was 146bp when run on a denaturing gel (Figure 3.15 lane 1), and was mostly composed of a discrete fragment of this size. Thus the majority of bands produced by the restriction digests are products of 146bp core DNA.

3.5.2 Nucleosome Mapping by Restriction Digestion of Core DNA

Four unique-cutting restriction enzymes, spaced evenly along the probe, were chosen to be able to differentiate between the positions of N1 and N2, previously mapped by ExoIII (Figure 3.13). In Figure 3.15B it is seen that *Ava*II should not cut core DNA protected by N2, and *Sma*I should not cut core DNA from N1. A test digest with each enzyme to be used, *Ava*II, *Acc*I, *Sty*I and *Sma*I on the full-length (240bp) nAB fragment, showed the products separated on a denaturing gel gave up to 4 bands (Figure 3.15A). This is due to the fact that, apart from *Sma*I, these enzymes do not leave blunt ends after cleavage. *Ava*II, *Acc*I, and *Sty*I leave 3, 2 and 4nt, 5′ overhangs respectively. The four bands are a result of the double stranded (ds) DNA being cut in two; each ds product runs as two separately labelled single strands on the

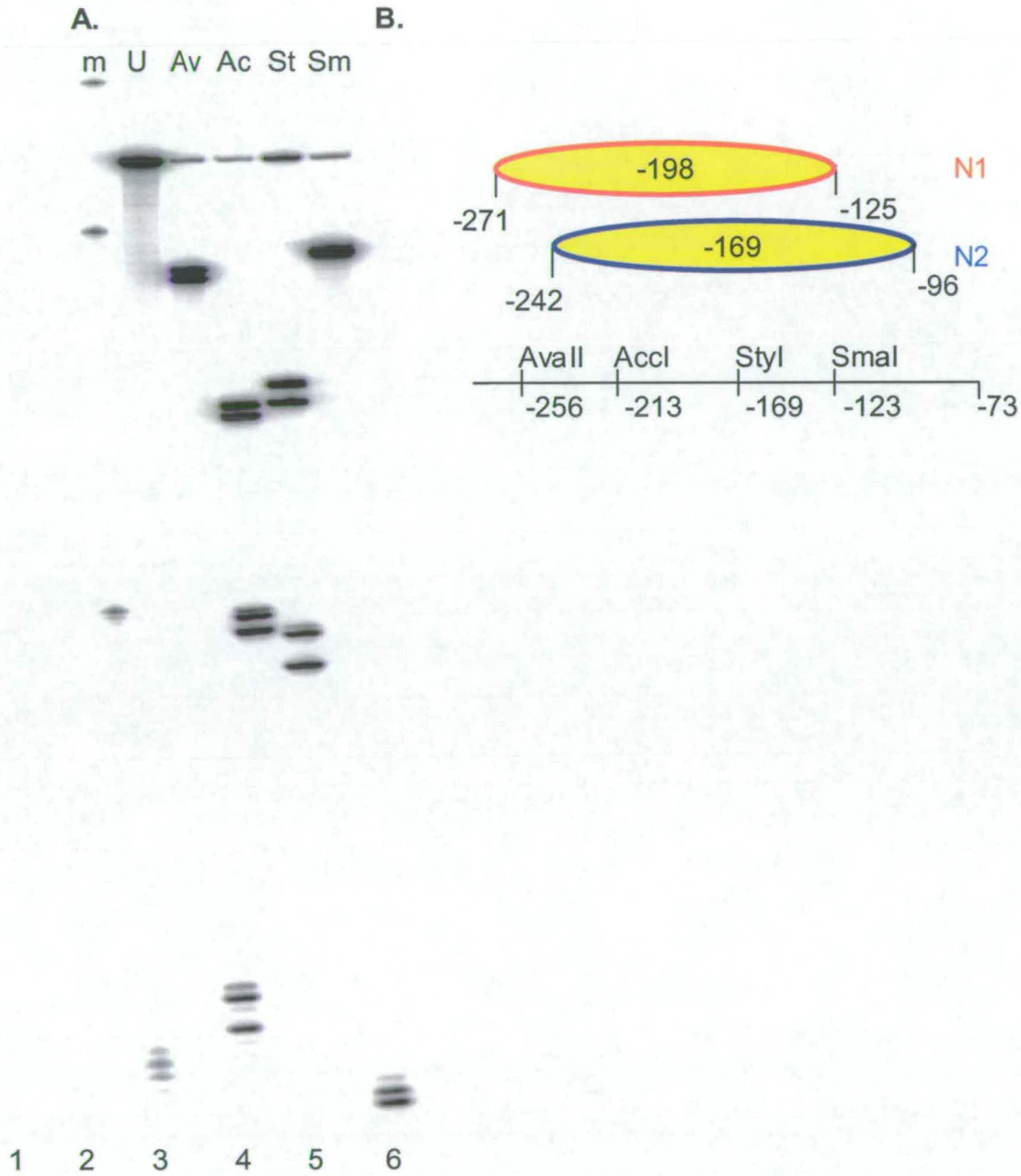


Figure 3.15 Restriction Test Digests.

A. Restriction digest of the full-length fragment nAB separated on a 10% denaturing polyacrylamide gel. Two bands for each digest product can clearly be seen. m, markers were 1kb ladder from fermentas; U, undigested; Av, *Ava*II; Ac, *Acc*I; St, *Sty*I; Sm, *Sma*I. **B** Location of restriction sequences on nAB, with a schematic diagram depicting N1 and N2 as mapped by ExoIII.

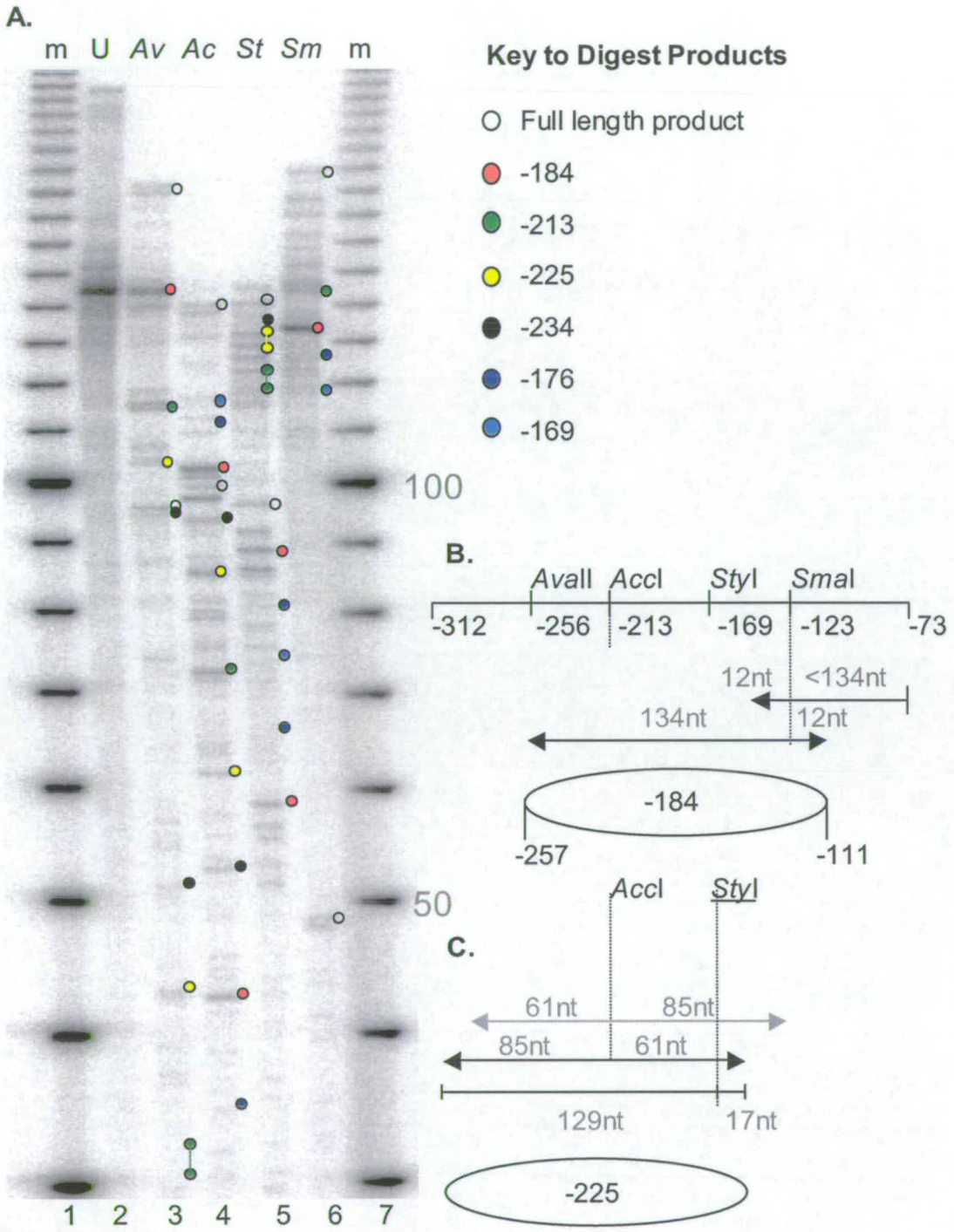
denaturing gel. These products are slightly different sizes due to the overhangs produced by the restriction enzymes. This could have been eliminated by end labelling. This technique was not used in this instance as MNase digestion would remove the signal, an alternative could have been to label purified core DNA.

Core particle DNA was digested with each restriction enzyme separately, and the products separated on a denaturing polyacrylamide gel (Figure 3.16). Although the band patterns formed are complex, the following examination indicates that patterns do emerge.

The *Sma*I digestion pattern (Figure 3.16, lane 5) is the simplest to interpret, containing no extra bands from overhangs to take into account. The location of *Sma*I towards the end of the fragment (Figure 3.15) also means it does not cut in every nucleosome isomer and consequently there is a significant amount of non-digested DNA at 146bp in the *Sma*I lane. The *Sma*I digest produces a very strong band of 134nt. The *Sma*I cleavage site itself is only 50nt from the downstream end of the full-length probe (Figure 3.16B), so this band must indicate a core particle boundary 134nt upstream from the *Sma*I site, i.e. at -257bp from the transcription start site. The corresponding 12nt fragment from *Sma*I to the downstream boundary at -111 is not resolved on this gel. These boundaries correspond to a positioned nucleosome with a dyad at -184bp. This position equates to the original *in vitro* mapped position at -183bp (Boa, 1999), and lies between the ExoIII mapped N1 position at -198 and N2 at -169. Bands in the *Acc*I and *Sty*I digests were also identified that correspond to a nucleosome at this position and are marked by red dots on the gel in Figure 3.16 and are listed in Table 3.2. The enzyme

Figure 3.16. Restriction Enzyme Digest of Core DNA.

A. Products of restriction enzyme digestion of core DNA were separated on a 10% denaturing polyacrylamide gel. Lane 2 contains core protected DNA from an unfractionated reconstitute on nAB. The same DNA was digested with unique cutting restriction enzymes, *Ava*II (lane 3), *Acc*I (lane 4), *Sty*I (lane 5) and *Sma*I (lane 6). Pairs of bands corresponding to each nucleosome position are colour coded as described in the key. Markers are 10bp ladder, band sizes in nt, calculated from the marker size are written in grey in order to be able to distinguish these from the position on the gene. Positions of the restriction enzyme sites on nAB are shown in **B**. Pairs of bands adding to 146bp, the length of the core protected DNA, have been identified (Table 3.1) and this data used to calculate nucleosome positions. The methodology used for calculation of the position of the nucleosome at –184 is displayed in **B** and for the –225 nucleosome position in **C**.


A.
m U Av Ac St Sm m
Key to Digest Products
Full length product
-184
-213
-225
-234
-176
-169
100
50
1 2 3 4 5 6 7
B.
AvaII AccI StyI SmaI
-312 -256 -213 -169 -123 -73
12nt <134nt
134nt 12nt
-184
-257 -111
C.
AccI StyI
61nt 85nt
85nt 61nt
129nt 17nt
-225

Calculated Position	*AvaII* -256	*AccI* *-213*	*StyI* *-169*	*SmaI* *-123*
-184	N/A	44 + 102	88 + 58	134 + 12
-213	30 + 116	73 + 73	117 + 29	N/A
-225	42 + 104	85 + 61	129 + 17	N/A
-234	95 + 51	94 + 52	138 + 8	N/A
-176	N/A	36 + 110	80 + 66	126 + 20
-169	N/A	29 + 117	73 + 73	119 + 27
Full length	56 + 184	99 + 141	143 + 97	189 + 54

Table 3.2 Pairs of Bands from the Restriction Digest of Core Protected DNA.

Calculated nucleosome positions are colour coded to match Figure 3.16. Pairs of digest products observed for each enzyme are listed under the relevant enzyme. Where there is more than one pair of bands due to overhangs from restriction enzymes the pair from the sense strand adding to 146nt is shown. N/A applies to nucleosome positions that are not cleaved by a particular restriction enzyme.

*Ava*II does not cut in the core produced by this nucleosome, as a result there is a strong band at 146bp in the *Ava*II lane indicating uncut core DNA.

Other nucleosome positions were similarly identified by allocation of pairs of bands adding up to 146bp. Potential positions were calculated using one restriction enzyme and then cross referenced with a second and third to work out an unambiguous position. An example is shown in Figure 3.16C. The position of the nucleosome at –225 (marked by yellow dots in Figure 3.16) was calculated initially from bands of 85 and 61nt in the *Acc*I digest (Figure 3.16A, lane 3). The *Acc*I data indicated either a nucleosome with boundaries at –274 and –128 (dyad at –201) or with boundaries at –298 and –152 (dyad at –225). *Sty*I (lane 4) digestion of a nucleosome at –201 should yield fragmentsof 105 and 41 nt. There is no evidence for these fragments in the gel. *Sty*I digestion of a nucleosome at –225 should yield fragments of 129 and 17nt. A fragment of 129nt is detected, indicating that the –225bp nucleosome is probably a correct position (Figure 3.16C). The corresponding 17bp fragment is not resolved on this gel. *Ava*II (lane 2) digest products of 42 and 104bp further confirm this (Table 3.2). The other nucleosome positions were assigned in a similar manner. All bands allocated to specific nucleosomes are marked by colour co-ordinated circles, and band sizes are recorded in Table 3.2.

Bands resulting from restriction digest of full length DNA were also assigned to products. Fragments of predicted sizes (recorded in Table 3.2) were discovered in all digests, these are marked as open circles in Figure 3.16.

3.6 Summary of Nucleosome Positions.

From the MNase / restriction mapping, nucleosomes positions were identified at –234, –225, –213, –184, –176 and –169bp from the transcription start site and by ExoIII digestion at –219, –198 and –169bp from the transcription start site. The ExoIII mapped nucleosomes can be linked to the isolated position isomer they were mapped from. Positions from restriction mapping must be linked to their corresponding isolated position isomer on the basis of the ExoIII positions or their predicted migration through a gel.

The strongest MNase / restriction mapped position at –184bp may correspond to the N1 ExoIII position at –198bp, as this is the strongest binding position in both, although these positions are 14bp apart.
The restriction mapped nucleosome position located at –169 mirrors the position for N2 derived from ExoIII analysis. The restriction mapped position located at –176 is also close to this figure and these may represent two closely positioned nucleosome binding sites in N2. These two nucleosome positions are likely to be represented in the *in vitro* map by the nucleosome A peak at –184 (and shoulder positions within this Figure 3.2).

The N3 nucleosome position isomer mapped by ExoIII lies at –219. The nucleosome positions mapped by restriction digests at –234, –225 and at –213 are therefore also likely to correspond to N3. These positions locate close to the upstream DNA end and thus are predicted to migrate as N3. The N3 nucleosome positions are represented in the *in vitro* map (Figure 3.2) by nucleosome B at –224, which is close to the MNase mapped position at –225. The main *in vitro* peak is flanked by two shoulders, which may represent alternative nucleosome position isomers, these appear to be spaced 10bp apart and could correspong to the MNase positions at –234 and –213bp.

3.6.1 Positioned Nucleosomes: Relevance to Stat5 Binding Sites.

Any of the positions mapped for N1 and N2 place the two outer Stat5 binding sites external to the nucleosome core (Figure 3.17) thus these both correspond to nucleosome position A. The nucleosomes at –213, –219, –225 and –234bp are likely to migrate as the isolated isomer N3. All of these positions place only one Stat5 binding site outside of the core DNA, therefore these positions correspond to nucleosome B. Some of the nucleosome positions place the Stat5 binding sites very close to the nucleosome boundary. The A1 binding site (Figure 3.17) is located between –279 and –271bp from transcription start. The upstream boundary of the nucleosome at –198 is at –272. This places the A1 Stat5 binding site right on the boundary of this nucleosome position. The strong nucleosome at –184 places 14bp between the A1 binding site and its boundary.

At the downstream end of nAB, the StM Stat5 binding site (Figure 3.17) lies between –94 and –86bp from the transcription start site. The downstream boundary of the nucleosome at –169 is located –96 bp from the transcription start site. This is two bp away from the StM Stat5 recognition sequence. The strong nucleosome position at –184 places 17bp between its boundary and the StM binding site.

3.6.2 Discrepancies Between MNase and ExoIII Maps.

The *in vitro* mapped position at –183 was mapped using MNase digestion, this position corresponds closely with the –184 position which was mapped by MNase and restriction digest mapping in this study, but is out by 15 bp from the N1 position mapped by ExoIII. It is unclear whether this difference

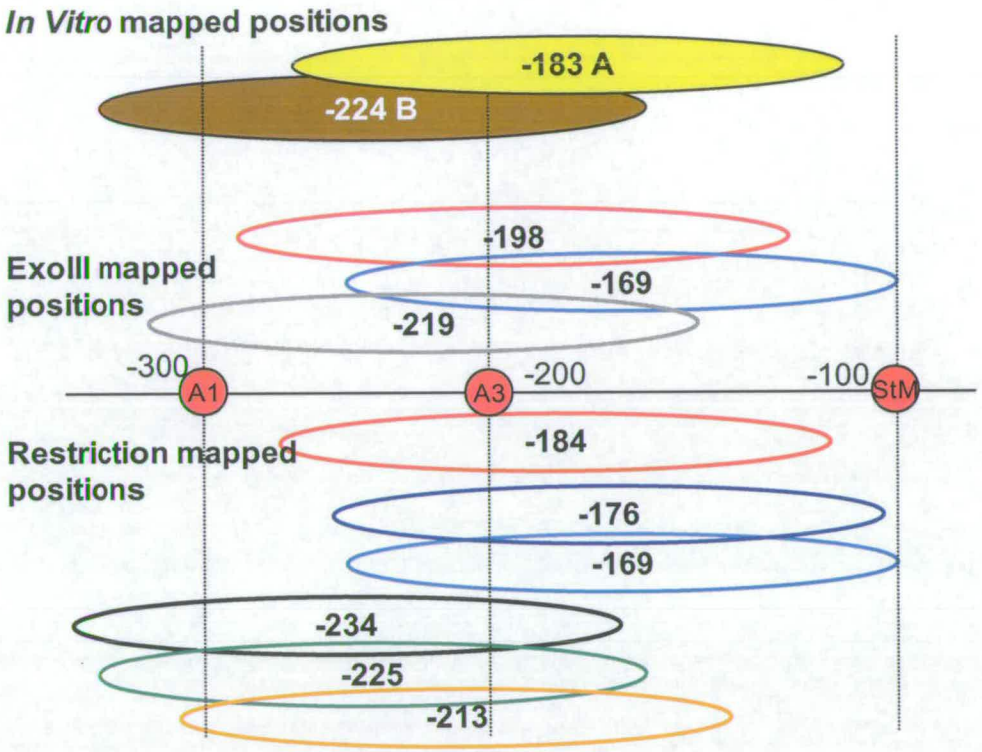


Figure 3.17 Summary of Mapped Nucleosome Positions.

The original *in vitro* mapped positions (A and B) are shown at the top, numbers indicate the dyad axis of the nucleosome core with respect to the transcription start site. Each nucleosome core is predicted to cover 146bp of DNA. N1 positions are indicated by red borders, and N2 by blue borders. N3 positions are coloured to match their corresponding bands in the MNase map. ExoIII maps positions N1 at –198 and N2 at –169. Restriction digest mapping maps N1 at –185 and N2 at –176/-169. N3 positions are –234, -225 and –213 bp from transcription start. Red circles mark positions of Stat5 binding sites on the promoter.

can be accounted for by the different mechanisms of the two enzymes, both regards the DNA sequence specificity and method of DNA attack.

Nucleosome positions on the beta globin gene have been mapped by both these methods (Kefalas et al., 1988; Yenidunya et al., 1994). The beta globin maps display a 3-4bp discrepancy between the positions mapped by MNase and ExoIII. Nucleosome positions over the *Saccharomyces cerevisiae* 5S rRNA gene have also been mapped using ExoIII, core restriction digests and DNaseI mapping. Nucleosome boundaries mapped by ExoIII and core restriction varied by up to 8nt (Buttinelli et al., 1993). Sequence specificity may be partially responsible for this, MNase cuts preferentially at A and T residues flanked by a G or a C residue (Cuatrecasas et al., 1967; Horz and Altenburger, 1981) while ExoIII has a tendency to pause at G residues (Linxweiler and Horz, 1982).

The sequence of the proximal region of the BLG promoter shows several strings of two, three and even up to five G or C nucleotides. Most of the strong ExoIII pause sites in the naked DNA lanes can be matched to runs of Gs. The (*Rsa*I) N2 band corresponds to a run of three G residues in the DNA. However there is not a detectable 10bp pattern following on from this point distinguishable from the contaminating N1 pattern. The N3 nucleosome at –292 mapped on (*Pvu*II) coincides with a run of Cs. ExoIII on (*Pvu*II) maps the lower strand, where these Cs will be Gs. A sequence specific pause site is observed for this run of Cs. None of the other nucleosome boundaries correspond to a run of Gs on the digested strand indicating that all pause sites observed are due to positioned nucleosomes.

The two positions mapped at nucleosome A by ExoIII (–198 and –169) are spaced outside the MNase positions (–184 and –176). This may be significant taken together with the fact that the ExoIII map of N1 covers slightly more DNA than the accepted 146bp. The sequence is GC rich (63% GC) which may affect the MNase digestion as this enzyme prefers to cut at AT, although the nucleosome core DNA produced was 146bp indicating that MNase digests right down to the nucleosome boundary.

Taken together these data provide a reliable estimate of the actual positions nucleosomes sit on these fragments and demonstrate that N1 and N2 locate two Stat5 binding sites in linker DNA, and N3 locates only one Stat5 binding site in linker DNA (Figure 3.17).

In summary, subtle sequence dependent effects within the β-lactoglobulin promoter result in two clusters of preferred nucleosome positions. One positioned between A3 and StM Stat5 binding sites leaving both these sites within the nucleosome linker region. The second position places one of these binding sites within a nucleosome. This arrangement may be significant in the transcriptional activity of the gene.

4 ANALYSIS OF STAT5 BINDING TO THE BLG PROMOTER.

The minimum 5′ flanking region of the BLG promoter required for maximal expression is 406bp (Whitelaw et al., 1992). This is termed the proximal promoter. Further dissection of the promoter revealed three binding sites for a mammary specific factor, MPBF (Watson et al., 1991), which was later shown to be a member of the Stat family of transcription factors and is now known as Stat5 (Burdon et al., 1994a; Wakao et al., 1994). These binding sites are named A1, A3 and StM and lie at –278, –210 and –93bp from the transcription start site respectively. Of these, the proximal site (StM) has the greatest affinity for Stat5. StM contains the full consensus sequence for Stat5 binding TTC(N)$_3$GAA whereas A1 and A3 both have single nucleotide changes from the consensus (Figure 4.1). A conserved binding site for Stat5 is present in the region around –90 bp from the transcription start site of the promoter of many milk protein genes, mutation of which abolished the hormone responsiveness of the promoters (Schmitt-Ney et al., 1992; Burdon et al., 1994b; Soulier et al., 1999; Schmitt-Ney et al., 1991; Demmer et al., 1995). The proximal promoter has also been shown to be responsible for the tissue specific expression of many milk protein genes (Lee et al., 1989; Whitelaw et al., 1992).

4.1 Stat5 Bandshifts.

The binding of Stat5 to these and similar sequences has been previously studied using the electrophoretic mobility shift assay (EMSA, or bandshift assay) (Watson et al., 1991; Schmitt-Ney et al., 1991; Burdon et al., 1994a; Burdon et al., 1994b; Soldaini et al., 2000). Recent work on the chromatin structure of the BLG promoter *in vivo* (Boa, 1999) has shown different nucleosome positions in tissues where BLG is active and inactive, which may

influence the ability of Stat5 to bind to the promoter (Chapter 3.1.1). This has led to our renewed interest in Stat5 binding to the BLG promoter.

In order to study further the interactions of Stat5 with the BLG promoter, bandshifts of oligonucleotides encompassing individual Stat5 binding sites from the BLG promoter by Stat5 from a mammary nuclear extract have been reproduced. The same bandshift interactions were also carried out on a larger fragment of the BLG promoter, nAB, which contains all three Stat5 binding sites see (3.2.1). Even though only one of the Stat5 binding sites on nAB fully obeys the recognised consensus, all three binding sites should be able to associate with Stat5. The binding of purified recombinant Stat5 (rStat5) to these probes is also studied in this chapter. Use of rStat5 facilitates comparison of the two forms of Stat5, Stat5A and Stat5B, and also the effects of the tetramerisation mutant W37A.

4.2 Stat5 Binding Sites Produce a Shift in Mammary Nuclear Extracts.

Previous bandshifts with Stat5 on the BLG and other milk protein promoters have used short oligonucleotide probes specific to individual binding sites with protein from nuclear extracts made from lactating mammary gland. These experiments were repeated to characterise the behaviour of Stat5 in the bandshift system described (2.8.3) before moving on to bandshifts with the longer probe nAB.

4.2.1 Nuclear Extract Bandshifts using Oligonucleotide Probes.

Oligonucleotide probes of the proximal Stat5 binding sites, StM and αs1 produce a specific shift in reactions with mouse, rabbit and ovine mammary gland nuclear extracts (Figure 4.2) and (Burdon et al., 1994a; Watson et al., 1991; Jolivet et al., 1996; Wakao et al., 1992; Pierre et al., 1994). Sequences

causing this shift pattern have been shown to be prolactin responsive in CHO and COS 7 cell lines (both with co-transfection of prolactin receptor) and in the mouse mammary cell line HC11 (Wakao et al., 1994; Soulier et al., 1999; Schmitt-Ney et al., 1991; Demmer et al., 1995; Burdon et al., 1994a; Burdon et al., 1994b). The presence of the proximal binding sites, specifically StM, is required for the prolactin responsiveness of the BLG promoter (Burdon et al., 1994a; Burdon et al., 1994b; Demmer et al., 1995).

Complimentary oligonucleotide pairs for Stat5 binding sites were designed for use in bandshift experiments. Stat5 binding sequences used are: StM, A1 and A3 from the BLG promoter and αs1, the proximal Stat5 binding site of the rabbit αS1-casein promoter; another strong Stat5 binding sequence, the affinity of which has been well characterised in bandshift experiments (Pierre et al., 1994). Point mutations designed to disrupt the consensus and predicted to prevent Stat5 binding were introduced to produce a second set (StMm, A1m, A3m). Sequences of the oligonucleotide probes that were used in bandshifts are shown in Figure 4.1.

Bandshift reactions with these oligonucleotide probes show two bands, of which the faster migrating (marked by the solid arrow) is stronger (Figure 4.2.). A similar pattern can be seen in many of the published Stat5-nuclear extract bandshifts (Meyer et al., 1997; Burdon et al., 1994a; Watson et al., 1991; Soldaini et al., 2000; Millot et al., 2001) and has been explained by the binding of multiple Stat5 dimers as a dimer-dimer complex, referred to as tetramerisation. This interaction usually involves cooperative binding to two weak Stat5 binding sites by interactions of the N-terminus of Stat proteins,

StM	GGGA**TTCCGGGAA**CCGCGTG
StMm	CAAGTGGGA**T**C**CCGGGA**TCCGCGTGGCTG
A1	GAAGTG**TTCCTGG**C**A**CTGGCAGCC
A1m	GACCCCGGAAGTG**T**A**CCTGG**C**A**CTGGCAG
A3	GGGGTC**TA**C**CAGGAA**CCGTCTAGGC
A3m	CTTCTGGGGTC**TA**C**CA**T**GG**TACCGTCTAGGCCC
αs1	GAGAA**TTCTTAGAA**TTTAAA

Figure 4.1 Sequences of Oligonucleotide Probes used in Bandshift Reactions.

Stat5 consensus binding sites are highlighted in bold and the mutations introduced to disrupt the consensus are underlined. The sense strand is shown here, the antisense is the exact reverse compliment. Each mutation is designed to introduce a restriction enzyme site as well as to destroy the Stat5 binding site. StMm contains both a *Sma*I (CCCGGG) and a *Bam*HI (GGATCC) site. A1m contains a *Rsa*I (GTAC) site and A3m both an *Nco*I (CCATGG) and a *Kpn*I (GGTACC) site. *Rsa*I will also cut at the *Kpn*I site of A3m.

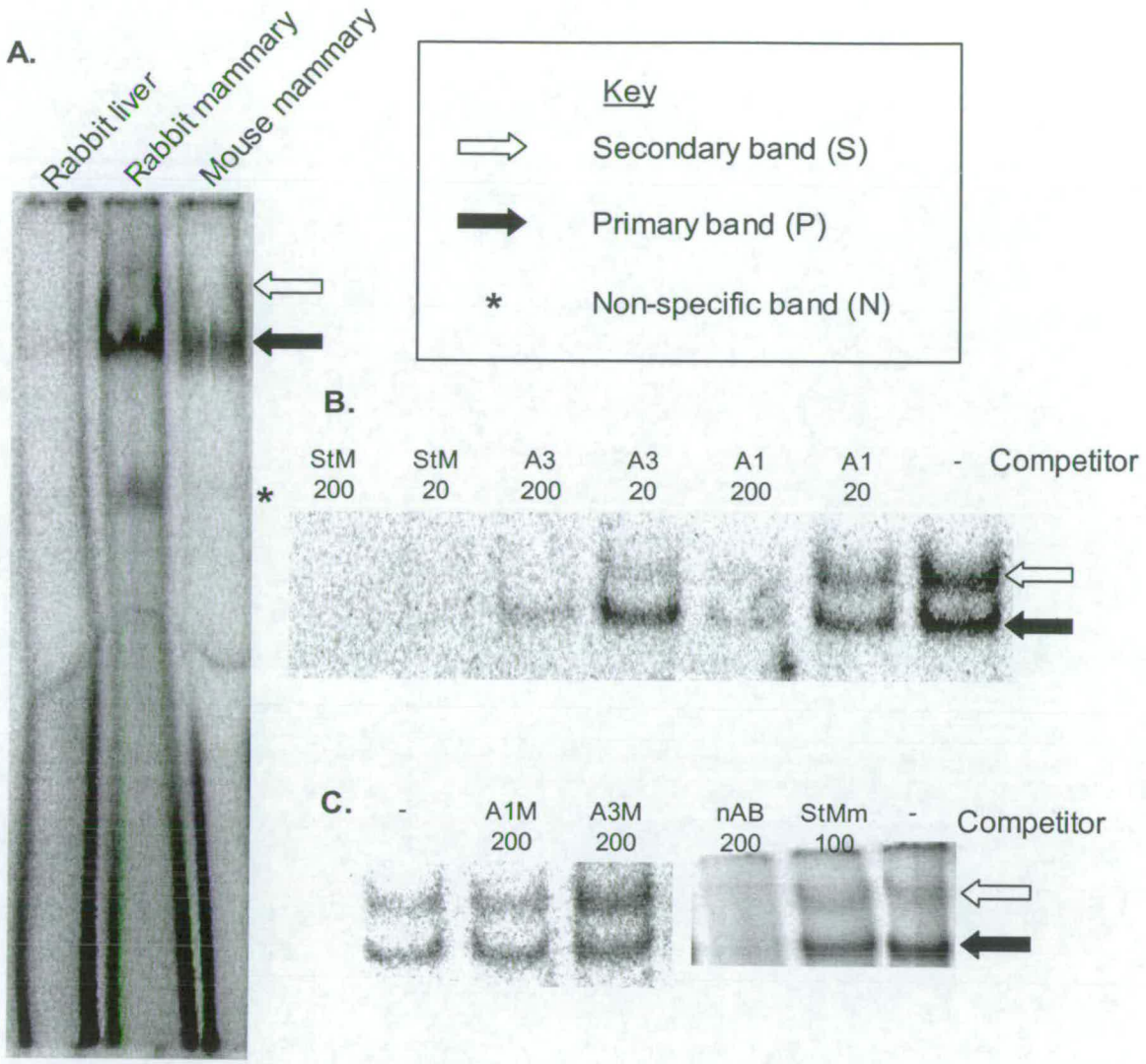


Figure 4.2 Bandshift Reactions with Oligonucleotide Probe.

A. Bandshifts with the αs1 oligonucleotide probe and nuclear extracts from rabbit liver, lactating rabbit mammary gland or mid lactating mouse mammary gland as indicated. Two bands are consistently seen with mammary but not with liver nuclear extracts, the faster migrating of which is the stronger. There is also a non-specific band marked by *. The same pattern is seen in Stat5 bandshifts with the StM probe in **B.** Cold competitor studies were carried out with the three Stat5 binding sites from the BLG promoter at indicated molar excesses. All are bandshifts with the StM oligonucleotide probe and mouse mammary nuclear extracts. Competition of Stat5 binding to StM with mutated binding sites and with the long probe nAB is studied in **C.** This demonstrates that the mutant probes do not bind Stat5. Lanes with no competitor are marked by a dash. Arrows mark shifted bands, as described in the key. The two halves of part **C** are from different gels, a non competed lane is included for each.

but can also occur on single Stat5 binding sites (Soldaini et al., 2000; Meyer et al., 1997). Tetramerisation can be blocked by mutation of tryptophan 37 to alanine (W37A). Both shifted oligonucleotide bands observed are competed equally by cold competitor. The secondary band could also be due to either other factors binding or differently phosphorylated forms of Stat5. This could also explain the differences in intensity of the secondary band in the shifts in Figure 4.2. The oligonucleotide probes were purified as double stranded so binding to contaminating single stranded probe cannot explain the two bands.

The αs1 probe is bandshifted by mouse and rabbit lactating mammary gland. Liver nuclear extract, which is traditionally used as a negative control for BLG expression also displays a weak bandshift activity with this probe (Figure 4.2a), the liver shift was also observed by (Burdon et al., 1994b). This result is not unexpected as active Stat5 is not confined to the mammary gland. It and other Stat species that will also bind to these probes are also found in many other tissues. The mammary gland shift is much stronger than that seen with liver, both reactions contain an identical amount of protein. A third band is also seen with the rabbit mammary, and the liver extract. This band is occasionally seen in the published Stat5 nuclear extract bandshifts, although no other binding sites are detected in the probe. No other shifted bands are observed with the oligonucleotide probes, so parts B and C of Figure 4.2 show only the two shifted bands.

Figures 4.2 B and C show bandshifts of the StM probe from the BLG promoter by mid lactating mouse mammary gland nuclear extract. The affinity of the various binding sites is tested by cold competitor studies,

where a known excess of unlabelled probe is added to the reaction after addition of the labelled probe. This technique is commonly used to compare the affinity of binding sites for proteins. Addition of a 20-fold molar excess of an unlabelled StM oligonucleotide completely competed the binding of Stat5 to the labelled probe and no shift is seen. When the cold competitor used is A1 and A3, although competition is clearly occurring, complete competition is not seen even at 200-fold excess (Figure 4.2 B), reflecting the different affinity Stat5 has for these compared to the consensus StM or αS1 binding sites. Competition with a 100 to 200-fold excess of each mutated Stat5 binding site (Figure 4.2 C) shows no difference from non-competed, demonstrating that the mutations created abolish Stat5 binding. The longer probe nAB also completely competes binding of Stat5 to StM at 200-fold excess. Lower concentrations of competitor were not tested.

4.2.2 Nuclear Extract Bandshifts using 240bp nAB Fragment.

Bandshift experiments with the same nuclear extract were repeated using the larger fragment of the promoter, nAB, which has been shown to bind to Stat5 by cold competitor studies (section 4.2.1). This probe contains three binding sites for Stat5, so may associate directly with up to three Stat5 dimers. Bandshifts with this probe and the same nuclear extracts used in oligonucleotide bandshifts show a smear between the free probe and the top of the gel rather than the individual discrete bands that are seen with the oligonucleotide probe (compare Figures 4.2A, and 4.3). This smear pattern is most likely due to the binding of multiple factors to the longer probe. There are consensus binding sites present for many transcription factors in the sequence of nAB including: NF-1, SP1, NF-kappaB, GR, YY1 and C/EBP alpha (Figure 1.8). There is no reproducible difference detected between the binding of probe nAB to lactating mammary nuclear extracts compared to.

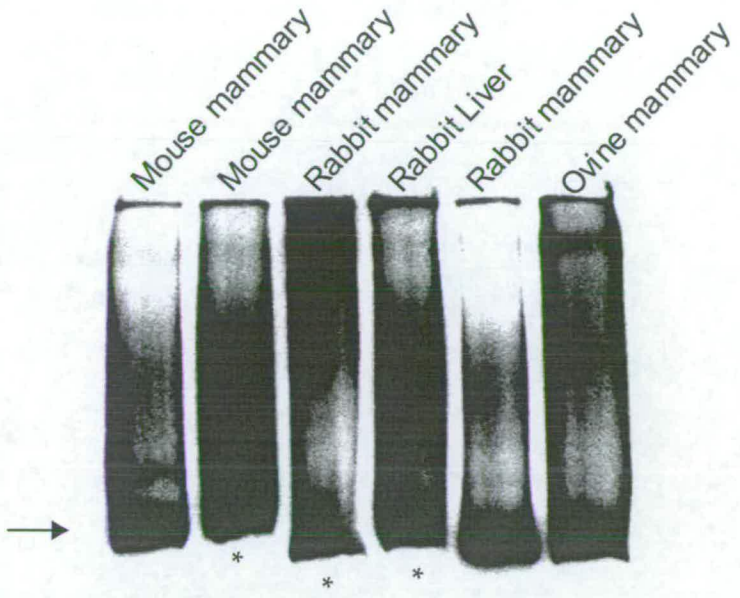


Figure 4.3 Nuclear Extract Bandshifts of nAB.

5μg nuclear extract from indicated sources was incubated with 5' end labelled nAB at room temperature and separated on a 5% polyacrylamide gel in TBE. An arrow at the bottom left indicates the free probe. The three nuclear extracts marked by * below the probe are the same as those used in Figure 4.2 A.

liver extracts that can be attributed to Stat5. For these reasons, Stat5 from a nuclear extract source will not be suitable for further studies on the interactions of Stat5 with the nAB probe and an alternative source must be found

4.3 Production of Recombinant Stat5.

To study further Stat5 binding to the BLG promoter, it is essential to be able to produce clear reproducible bandshifts using the nAB probe. The most straightforward method to eliminate binding of other transcription factors is to use purified Stat5. Previous attempts to purify Stat5 from tissue have produced an amount just greater than 0.01mg with a specific activity of 3,000,000u/mg (one unit being defined as the amount which causes the retardation of 1fmol labelled DNA probe in a bandshift assay,) from 800g of rat lactating mammary gland (Wakao et al., 1992); or less than 0.01mg (with a specific activity of >15,000,000 units per mg,) from 6kg of ovine lactating mammary tissue (Wakao et al., 1994). These purifications involved multiple stages of separation using Bio rex 70, DNA-sepharose and a specific DNA affinity column.

Other groups have used a baculovirus expression system (BEVS), to produce recombinant Stat proteins, including Stat5 containing a 6-His tag engineered into the N-terminus for ease of purification (John et al., 1999). Use of recombinant Stat5 also allows the introduction of mutations into the Stat5 molecule, such as the tetramerisation mutant W37A, and separate analysis of the two forms of Stat5, Stat5a and Stat5b.

4.3.1 Baculovirus Expression System.

Bandshifts have previously been carried out using Stat5 produced in a baculovirus expression system (BEVS), so this approach was also chosen to produce recombinant Stat5 for use in these studies. Use of the BEVS has many advantages over other methods of producing recombinant protein. Proteins produced by over-expression in bacterial systems are often not correctly folded or post-translationally modified. Insect cells however process and modify proteins in a similar manner to mammalian cells (Smith et al., 1983; Miyamoto et al., 1985). It has been reported that Stat proteins produced in the BEVS are phosphorylated on the conserved tyrosine that is normally only modified *in vivo* in the activated form of Stat (section 4.3.6 and (Soldaini et al., 2000; John et al., 1999)). Phosphorylation on this tyrosine is essential for Stat DNA binding activity. This phenomenon eliminates the need to activate Stat5 *in vitro* before it will bind DNA in a bandshift reaction.

A recombinant Stat5 baculovirus was produced by cloning Stat5 into an altered virus genome under the control of the polyhedrin promoter. The polyhedrin promoter in a wild type virus would normally drive expression of polyhedrin, the viral coat protein, and as such it is a strong promoter (Smith et al., 1983; Pennock et al., 1984). High yields of recombinant protein are produced under the correct conditions. Typical yields are much higher than expression in a mammalian culture system could produce, and are capable of being greater than 50% of total cell protein (Miyamoto et al., 1985; Maeda et al., 1985). Host protein synthesis is shut down during infection so most of the cell's resources are directed towards the production of viral proteins.

4.3.2 Cloning Stat5 into pFastBac Vectors.

Human Stat5 was cloned into a baculovirus expression vector using the Bac-to-Bac kit from Invitrogen. Stat5a (pHuStat5a~6HisBacPAK8), Stat5b (pHuStat5b~6HisVL1392), Stat5aW37A (pCi-Stat5aW37A) and Stat5bW37A (pCi-Stat5bW37A) DNA sequences were a gift from W J Leonard. pFastBac1 and pFastBacHT vectors are from Invitrogen. Wild type Stat5a and 5b were cloned into pFastBac1, and Stat5aW37A and 5bW37A were cloned into pFastBacHT as described below. Wild type Stat5a and 5b contain a 6-His tag engineered after the third amino acid; this is used to purify the recombinant protein.

Stat5a cDNA (accession number u43185, (Schindler et al., 1995; John et al., 1999)) was removed from pHuStat5a~6HisBacPAK8 as a *Sac*I - *Sma*I fragment and cloned into the Multiple cloning site (MCS) of pFastbac1 (that had previously been linearised with *Nco*I, the ends blunted with T4 polymerase, then digested with *Sac*I) to produce pFastbac1Stat5ax. However this method inadvertently introduced a second ATG site between the polyhedrin promoter of pFastbac1 and the Stat5 transcription start site ATG codon, which was not in frame with the protein (Figure 4.4A, the extra ATG is shown in grey). In order to correct this, a site directed mutagenesis PCR was carried out to introduce a *Bss*HII restriction site between the two ATG start codons. This amplified a fragment of 400bp, which includes a unique *Nde*I restriction site internal to the Stat5a gene. There are now two *Bss*HII sites, flanking the extra ATG. This PCR product was cloned into pGEM—T (Promega) to give pGEM-T5ax. Digestion to completion with *Bss*HII and *Nde*I results in a 330bp fragment, encompassing the correct transcription start (Figure 4.4B). A short fragment containing the extra ATG, was discarded. The 330bp fragment was cloned into the original pFastBac1Stat5ax that had

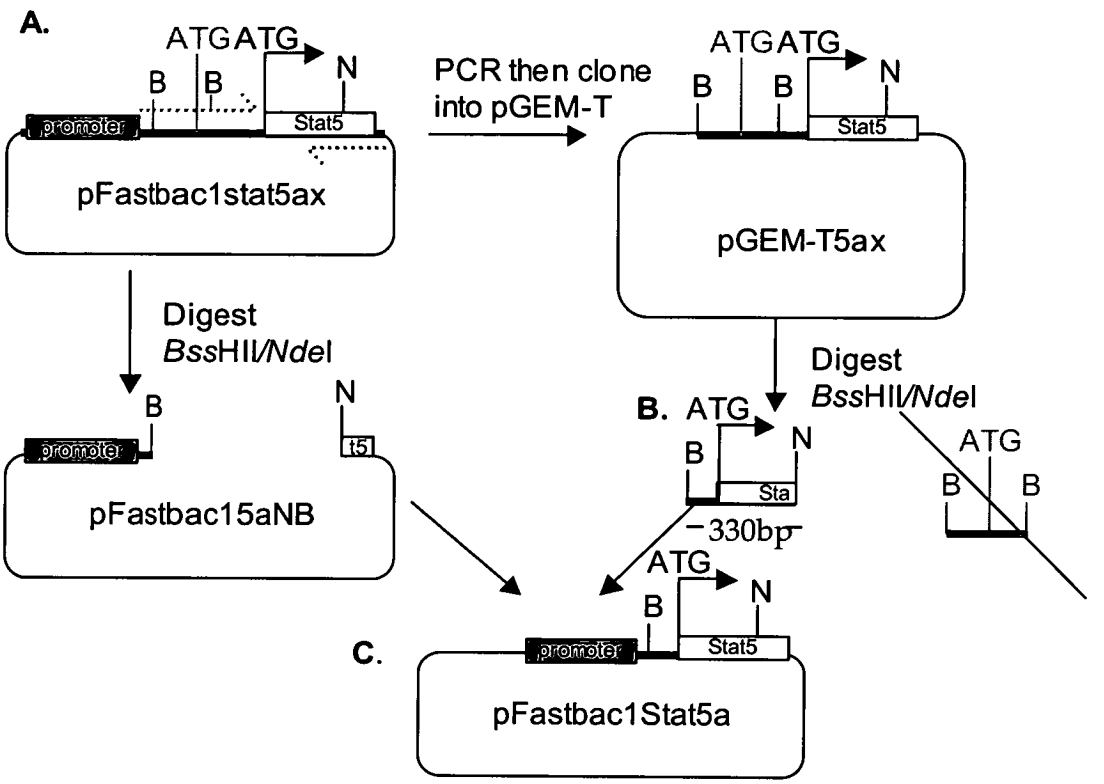


Figure 4.4. Strategy for Removal of the Extra ATG in pFastbac1Stat5ax.

Schematic diagram describing the method used to remove the extra ATG from pFastbactStat5ax. **A**. A site directed mutagenesis PCR was carried out to insert a *Bss*HII recognition site between the two ATG sites. The product was cloned into pGEM-T **B**. An *Nde*I/*Bss*HII digest removes a fragment containing only the correct ATG and the beginning of the gene. This is cloned into the original plasmid opened with *Nde*I/*Bss*HII to replace the sequence containing the incorrect ATG, **C.** resulting in Stat5 cloned correctly into pFastbac1.
N = *Nde*I, B = *Bss*HII.

been digested with *BssH*II and *Nde*I and purified to remove the incorrect ATG sequence (pFastbac15aNB, Figure 4.4C) to produce pFastbac1Stat5a. The correct sequence was verified by sequencing.

Stat5b (accession number u47686 (Lin et al., 1996)) was removed from pHuStat5b~6HisVL1392 as an *Eag*I / *Xba*I fragment and cloned into pFastbac1opened with *Not*I / *Xba*I. The correct insertion was verified by restriction enzyme analysis and sequencing.

The Stat5aW37A and Stat5bW37A DNA sequences do not contain an internal 6-His tag. In order to purify them they were cloned into a vector, pFastbacHT, which is designed to add a 6-His tag to the N terminus of the protein. This 6-His tag can be removed by AcTEV protease digestion of a 7 amino acid recognition site situated between the tag and Stat5. Three vectors HTA, HTB and HTC are available, each has the 6-His tag in a different reading frame. Stat5aW37A was cloned into pFastbacHTB that had been opened with *Nco*I and *Not*I. *Nco*I was chosen as its position at the start of the gene brought it as close as possible to the 6-His tag in the pFastHT vector, minimising the inserted amino acids at the N terminus to two (after removal of the tag). *Nco*I cuts three times internal to the Stat5a coding sequence so Stat5aW37A was cloned into pFastHTB by three separate steps.

The cloning of Stat5aW37A from pCi~Stat5aW37A into pFastHT is outlined in Figure 4.5. The Stat5aW37A gene was digested into three fragments by a combination of *Eco*RI, *Nco*I and *Not*I digests (Figure 4.5A and B). *Eco*RI was used to separate two of the *Nco*I sites. The first step was to clone the central

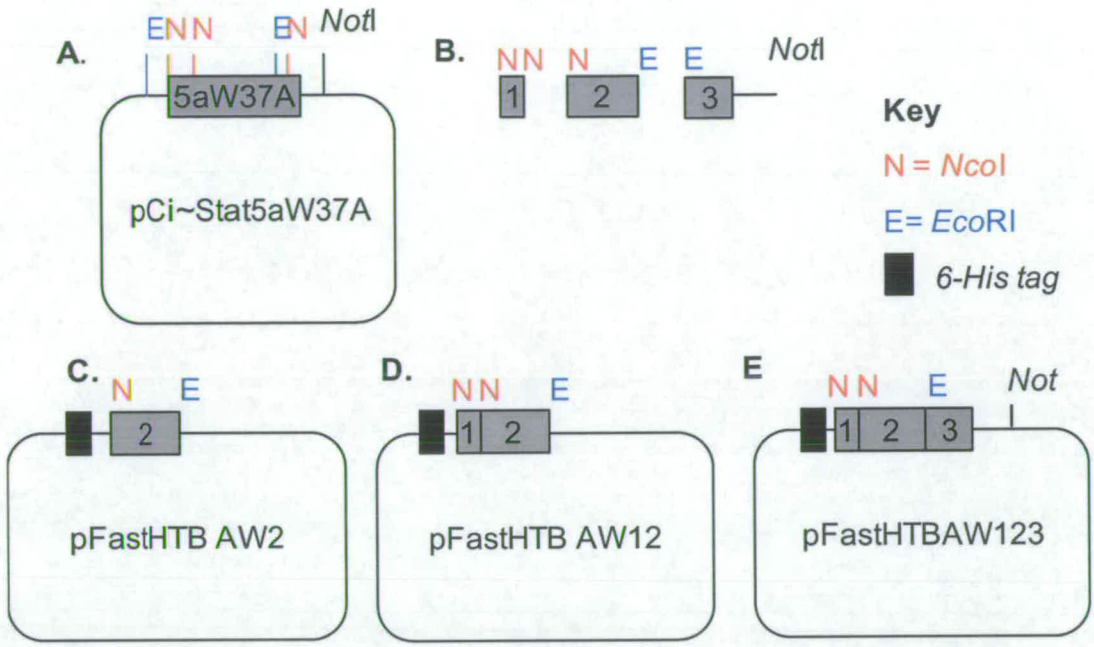


Figure 4.5 Cloning Stat5W37A into pFastbacHT.

Schematic diagram of the method used to clone Stat5aW37A into pFastHTB. The restriction enzyme sites in pCi~Stat5aW37A are shown in **A**. Restriction digest products of 5aW37A are shown in **B** (grey boxes indicating fragments 1 2 and 3). The three cloning stages into pFastHTB are shown in **C, D** and **E.** Cloning of Stat5bW37A into pFastHT was by a similar method not detailed here.

part of the gene (fragment 2), from *Nco*I to *Eco*RI into pFastHTB to create pFastHTB-AW2 (Figure 4.5C). The second step was to clone the most 5' 125bp of Stat5aW37A (fragment 1) as an *Nco*I-*Nco*I fragment into pFastHTB-AW2 to give pFastHTB-AW12 (Figure 4.5D) this product was verified by sequencing, as the insertion was too small to be sure of the correct orientation and copy number by restriction analysis. Finally, in the third step, the 3' end of the gene (fragment 3) was inserted as an *Eco*RI to *Not*I fragment into pFastHTB-AW12 to give pFastHTB-AW123 (Figure 3.5E).Stat5b was cloned into pFastHT by a similar method, but as Stat5b contains only two *Nco*I recognition sites there was one less step. Effectively the two cloning steps were as steps one and three from the Stat5a cloning. The final products were verified by restriction analysis and sequencing.

4.3.3 Production of Virus.

Recombinant baculovirus was produced using the Bac-to-Bac baculovirus expression system from Invitrogen. In this system the recombinant viral DNA is produced in *E.coli* DH10Bac cells, rather than in insect cells as with the traditional BEVS. It also makes use of blue white screening, which saves time and effort and eliminates the need for repeated rounds of virus purification to obtain a pure recombinant virus as only the correct recombinant DNA is introduced into insect cells.

The pFastbac vectors contain two Tn7 transposon attachment elements that flank the MCS, and a gene encoding resistance to gentamycin. pFastbac vectors were transformed into DH10Bac competent cells, which contain a baculovirus shuttle vector (bacmid) and a helper vector. The bacmid includes a mini-attTn7 target site and a kanamycin resistance gene. The helper vector encodes for tetracycline resistance and Tn7 transposase. The

transposase inserts the Tn7 target sequence from a pFast vector, including any sequences that were cloned into the MCS, into the Tn7 attachment site on the bacmid by a transposition reaction. The Stat5 gene, cloned into the MCS of pFastBac1, and gentamycin resistance are now found as recombinant virus DNA in the bacmid. Insertion of the Tn7 sites disrupts the LacZα gene allowing blue white screening for recombinant bacmid DNA. Growth in the presence of the three antibiotics, tetracycline, kanamycin and gentamicin also selects for colonies that have successfully integrated. To screen for correct insertion, minipreps of the high molecular weight DNA were amplified by a PCR reaction with M13 primers. These amplify the DNA sequence between the Tn7 attachment sights. Primer sequences are:

M13F **GTTTTCCCAGTCACGAC**

M13R **CAGGAAACAGCTATGAC**

An empty bacmid results in a product of 300bp, and Stat5a cloned into pFastbac1 generates a product of 5.5kb. This is in agreement with the inserted Stat5a DNA length of 3.2kb plus 2.3kb from the pFastbac vector. Figure 4.6 shows a typical experiment. Lane 5 is the 300bp product from an empty bacmid vector, and lane 2 is the 2.3kb product from an insertion of an empty pFastbac1 vector. Lanes 4 and 6 may be an incorrect Stat5 insertion and lanes 7 and 8 are the products from a correctly inserted Stat5a recombinant baculovirus.

Approximately 1ng of each Stat5 recombinant baculovirus DNA was transfected into mid-log SF9 cells using Cellfectin reagent (Invitrogen). After 72 hours, the cells start to produce virus. The medium was collected and clarified before being used to infect fresh mid log cells to amplify the stock. This first viral collection is termed the P1 stock and the second the P2 stock.

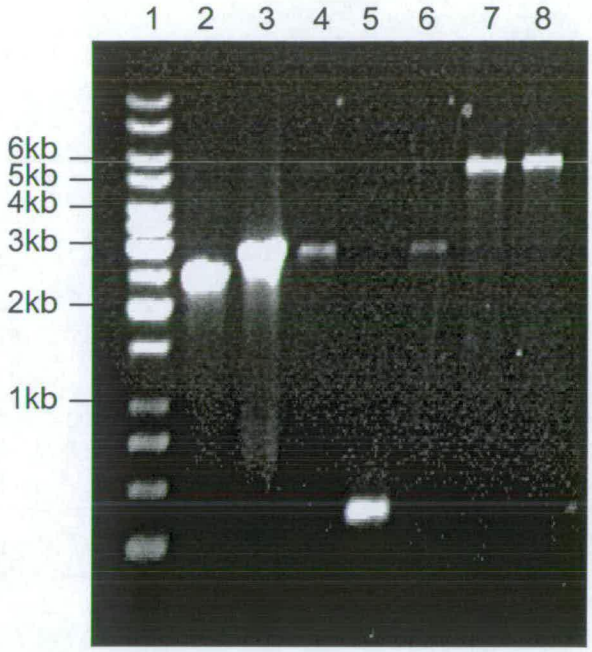


Figure 4.6. Products of M13 PCR Reaction.

Verification of the correct insertion of Stat5 genes into the bacmid was carried out by a PCR reaction. Recombinant bacmid DNA preps were amplified using M13-Forward and M13-Reverse primers. The PCR on an empty bacmid (lane 5) gives a product of 300bp. Lane 2 contains the product from an empty pFastbac1 insertion and lane 3 from a pFastHTCAT control insertion. Stat5a recombinant bacmid DNA reactions are lanes 4 and 6 to 8. Of these lanes 7 and 8 are the expected size of 5.5kb. Lane 1 is 1kb ladder.

4.3.4 Virus Titer.

Virus titers of the P2 virus stocks were calculated in plaque forming units (pfu) per ml using the BacPAK Baculovirus Rapid Titer Kit (Clontech). P2 stocks were typically in the range 0.3 to 3 $x10^8$ pfu/ml in a volume of 2ml, a second amplification step to obtain a larger volume of stock was termed P3.

Every recombinant baculovirus produced will behave slightly differently as regards progression of the infection and maximum production of recombinant protein. The multiple of infection (MOI) and time of harvesting employed can be varied to maximise the yield. Previous studies expressing Stat proteins in a baculovirus system (Schindler et al., 1995; John et al., 1999) have used MOIs of 5 pfu per cell and harvested cells at between 66 and 72 hours post infection. This was taken as a starting point in the determination of the optimal conditions for production of Stat5 in this baculovirus expression system.

4.4 Analysis of Protein Expression.

Test baculovirus expression studies were carried out to determine the optimal MOI with which to infect SF9 cells, and the best time point to harvest the cells in order to maximise Stat5 yield.

$1x10^6$ SF9 cells per well of a 12 well plate were infected at MOIs of 1, 2, 5, 10 and 20 pfu/ml, and harvested at either 24, 48, 66, 72 or 96 hours post infection. Equal amounts of whole cell extracts, were separated on NuPage protein gels, and Western blotted with an anti-Stat5 antibody that recognises both Stat5a and 5b (Zymed 33-5900) to determine the total yield of Stat5. The lower half of the membrane was probed with an anti-SF9 cell line antibody

(Biodesign K47012R) as a loading control; this antibody picks up many bands but one discrete band migrating at about 37kDA was used to normalise Western blots to any differences in protein concentration.

At low MOIs (of one or two) Stat5 is only detected after 72 hours (data not shown). No Stat5 was detected at 24 hours at any MOI. By 48 hours Stat5 begins to be detected at higher MOIs.

These results show that, contrary to previous studies, maximal levels of Stat5 protein were produced at 96 hours post infection, and with a relatively high MOI of 20 pfu/cell (Figure 4.7A, middle panel and 4.7B light blue bars). However, it is not just the yield of Stat5 that is important. As I intend to utilise the intrinsic phosphorylating activity of SF9 cells to produce phosphorylated, and therefore active, Stat5 the level of this also must be assessed.

4.4.1 Phosphorylation Status.

Recombinant Stat protein produced in a baculovirus expression system has been reported to have DNA binding activity and to be phosphorylated on the conserved tyrosine residue (Y694 in Stat5a, Y699 in Stat5b) (Lin et al., 1996; Soldaini et al., 2000) that is modified on activation in the JAK/Stat pathway, without the need for additional stimulus. A mechanism native to the SF9 cell line must be causing this phosphorylation, a Stat family protein has been found in insect cells (Yan et al., 1996; Hou et al., 1996) indicating the existence of pathways capable of phosphorylating Stat proteins.

In order to characterise the intrinsic phosphorylating behaviour of the SF9 cells, the blots described in section 4.3.5 and Figure 4.7 were stripped and re-

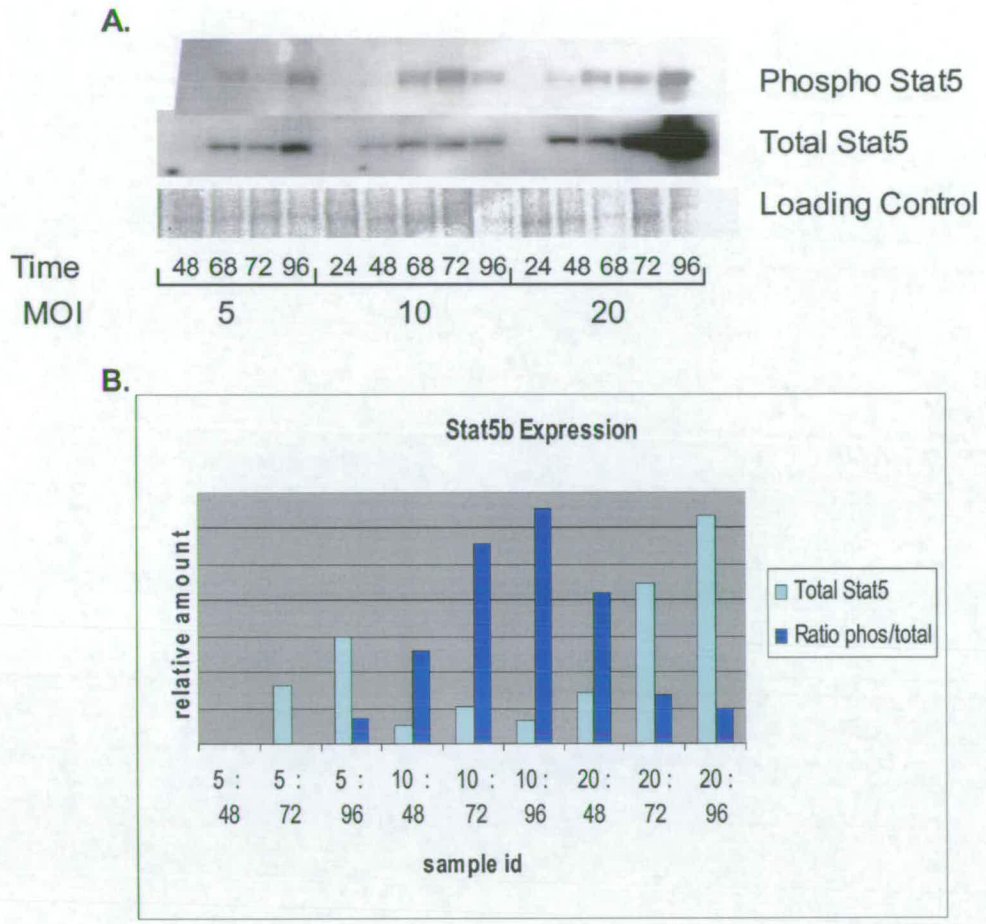


Figure 4.7 Stat5b Expression Tests.

$1x10^6$ SF9 cells were infected with Stat5b at indicated MOIs and the total protein harvested at the indicated times. Aliquots were separated on NuPage protein gels and transferred to nitrocellulose membrane. Membranes were probed with antibodies against pan-Stat5 (central panel) and pan-phospho-Stat5 (top panel). The western blots are shown in part **A**. Density of signal at each time point was normalised to the protein loading. The total amount of Stat5 plus the ratio of phosphorylated Stat5 over total Stat5 for each sample (MOI : time of harvest)are shown in the graph in part **B**.

probed with an anti-phospho-tyrosine-694 Stat5a/b antibody (Zymed 33-6000) to determine the optimal conditions for maximum yield of the phosphorylated protein.

The levels of phosphorylated Stat5 showed a different pattern to total Stat5 (Figure 4.7 compare the top and middle panels). When this data was analysed it was revealed that the proportion of phosphorylated, and therefore active, Stat5 that is produced changes with time (Figure 4.7B). Although more Stat5 is produced after 96 hours, the level of phosphorylated Stat5 remains more constant meaning less of the Stat5 produced will be capable of binding DNA in a bandshift reaction. It may be that proteins involved in the phosphorylating activity are down-regulated with the rest of the host proteins during the course of the infection, or that the rate of production of Stat5 may be greater than the rate at which it can be phosphorylated.

The relative amounts of phosphorylated Stat5 over total Stat5 are shown in Figure 4.7B (blue bars). When this is compared to the relative levels of total Stat5 (light blue bars) the time point chosen as having the greatest proportion of phosphorylated Stat5 is 72 hours post infection, with a MOI of 10 pfu/cell. These were the conditions used for the large scale production of Stat5. The time point of 96 hours postinfection at a MOI of 10 was not chosen, although it appears to contain a high proportion of phosphorylated Stat5, as 72h with a MOI of 10 actually produces a higher level of phosphorylated protein.

4.4.2 Nickel Column Purification.

The Stat5 described in this chapter has been engineered to contain a 6-His tag at the amino terminus. Proteins tagged in this way can be easily isolated

using nickel (Ni^{2+}) affinity chromatography. Amino acids, particularly histidine and cysteine have affinity for divalent cations such as nickel, zinc and cobalt (Porath et al., 1975). Nickel is generally considered the best metal to use, and sufficient purification was achieved with this that it was thought unnecessary to try to optimise the system with an alternative metal. The run of six histidine residues in the 6-His tag has a higher affinity for Ni^{2+} than most proteins; this property is made use of to purify Stat5 from the total protein. Imidazole, as an electron donor, competes with histidine and other amino acids for interactions with the metal ions (Porath, 1992). Ni^{2+} charged HiTrap Chelating HP columns (Amersham Biosciences) were used to purify rStat5 from whole cell extracts of baculovirus infected SF9 cells. Cell extracts were loaded in a binding buffer containing sufficient imidazole concentration to efficiently compete the binding of as much of the total protein as possible, while allowing all the tagged rStat5 to remain bound. The elution buffer was chosen with an imidazole concentration sufficient to ensure that all tagged rStat5 eluted in one step.

Optimum imidazole concentrations were selected in test purifications. A whole cell extract of SF9 cells infected with a 6-His tagged Stat5a baculovirus in phosphate buffer with 10mM imidazole was loaded onto a 1ml column. Five ml aliquots of phosphate buffer containing increasing imidazole concentrations were added between washes of phosphate buffer with 10mM imidazole. The eluates were collected and separated on a protein gel. The imidazole concentration immediately before Stat5 is first eluted is the optimal binding buffer. Stat5 first appears with elution at 40mM imidazole, (data not shown) so the binding buffer used was 30mM imidazole. Stat5 was eluted in 300mM imidazole.

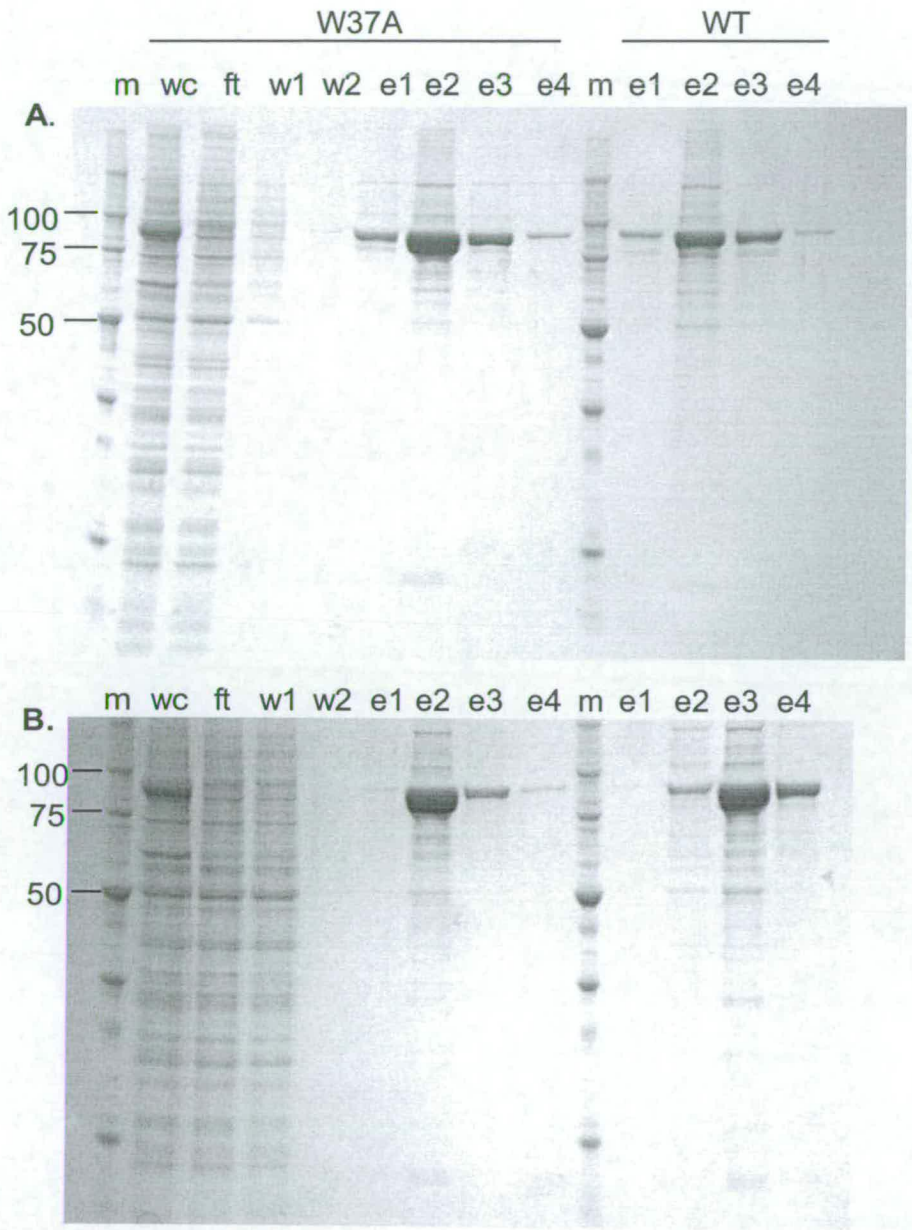


Figure 4.8. Purified rStat5.

Coomassie stained 12% BisTris NuPAGE gels showing the purification of 6-His Stat5 on a nickel charged column. **A** is Stat5a and **B**. is Stat5b. **m** are Precision Protein Standards (Bio-Rad), **wc** is whole cell extract, **ft** is the column flowthrough, **w** are the wash steps and **e** represents the elution steps (all at 300mM imidazole). The whole cell, flowthrough etc is shown for both W37A mutants (between the two marker lanes) and only the elution steps for the WTStat5 (to the right of the central marker lane). Stat5a WT and W37A and Stat5b W37A were eluted in e2 and Stat5b WT was eluted in e3.

Stat5 purified under these conditions is >95% pure (Figure 4.8). The rStat5 band is clearly seen in the whole cell lane at about 90kda, and is not present in either the flow through or the washes. There is also a second band at approximately 60kda present in the whole cell that is reduced in the flowthrough and in the first wash step but this band is not enriched in the purification steps. Stat5 can be observed eluting usually in the second or third fraction. Typical yields of rStat5 were between 3 and 6.3mg rStat5 per 10^9 SF9 cells.

4.4.3 Cleavage of 6-His Tag.

The 6-His tag on the Stat5 W37A mutants was designed to be removed by AcTEV protease. A seven amino acid recognition site (Gln-Asn-Leu-Tyr-Phe-Gln-Gly) is inserted between the 6-His tag and the Stat5 transcription start by the pFastHTB vector. The AcTEV used (from Invitrogen) also contains a 6-His tag so it can be removed from the sample after cleavage. Cleavage of the 6-His tag can be followed by staining the gel with InVision™ His gel stain from Invitrogen. This stains proteins containing a poly histidine tag, causing them to fluoresce when excited by UV light.

Test digests of Stat5aW37A and Stat5bW37A with the AcTEV protease (Figure 4.9) show that in these gels Stat5 migrates as a group of bands, in contrast to Figure 4.8 where only a single band is seen. The fastest migrating band appears to be the most resistant to AcTEV digestion, as only this and one other band remains in the Stat5bW37A digest after digestion for one hour, no other bands are seen in the Stat5aW37A digest at this time point. The slower migrating bands that disappear almost immediately after AcTEV cleavage are also not observed in the Coomassie stain of Stat5bW37A suggesting that total degradation of these species has occurred rather than

Figure 4.9 AcTEV Digests of Stat5 W37A.

rStat5a W37A (left hand panels) and rStat5b W37A (right hand panels) was digested for the indicated time periods with AcTEV to remove the 6-His tag, then separated on NuPAGE gels under reducing conditions. Gels were stained with InVision His tag gel stain to display 6-His tagged proteins (upper panels) then coomassie stained to display the total protein.

removal of the 6-His tag. In fact the total protein stain mirrors the gradual disappearance in the 6-His stain of the fast band that is resistant to degradation. In the Stat5aW37A digests the 6-His tagged bands disappear faster than they do in Stat5bW37A, only a trace is seen after 4 hours whereas for Stat5b a band is still present after 7 hours.

In both digests, bands are seen at 80kda and below with the Coomassie stain that do not stain for the His tag, indicating cleavage of the 6-His tags. A band at 60kDa appears at one hour of AcTEV digestion of Stat5aW37A, and is also observed at later stages of digestion of Stat5bW37A. These bands are too small to be full length Stat5. Bands of between 20 and 30 kda appear in digested samples with the 6-His stained gel but are not seen with the Coomassie stain, these may represent the cleaved part of Stat5 (perhaps the other part of the 80kda band) or the AcTEV. Both of these contain a 6-His tag. Similar patterns were seen under other digestion conditions (data not shown). It appears that AcTEV digestion degrades Stat5 and that removal of the 6-His tag from the Stat5W37A is not a viable option. A search in the Stat5 sequence for a motif resembling the AcTEV recognition site did not pull up any possible cleavage sites. Stat5bW37A in general appears to be more resistant to degradation than Stat5aW37A.

4.5 Bandshifts using rStat5.

The 6-His tagged rStat5 described in section 4.3 was used in bandshifts of both the oligonucleotide probe StM and the 240bp probe nAB. Binding patterns for each rStat5 on each probe are described in the following sections.

4.5.1 Recombinant Stat5 Bandshifts of StM

The probe StM should only be able to associate directly with one Stat5 dimer, as it contains only a single binding site for Stat5. Bandshifts using this probe with rStat5 show multiple bands (Figure 4.10 compare lane 2 (nuclear extract) with lanes 3 – 6 (rStat5)). It has been suggested that this may be due to the binding of either alternatively phosphorylated forms of Stat5 or truncated (perhaps degraded) Stat5. Phosphorylated Stat5 has been known to migrate at a different level from non-phosphorylated (Vinkemeier et al., 1996). Alternatively it must be noted that these recombinant Stat5 proteins contain a 6-His tag at their N-terminus (plus other sequences for W37A) which may be a factor in the migration of DNA bound complexes.

Previously, bandshifts with nuclear extracts have shown a pair of bands specific to Stat5 (Figure 4.2). The slower migrating secondary band, labelled S in Figure 4.2, might not show in the nuclear extract shift in Figure 4.10, as it is generally much weaker than the primary band. In fact with different nuclear extracts the relative intensity of these bands vary enormously. The whole gel including the free probe is shown in Figure 4.10 and no non-specific bands are seen.

Stat5 produced in the baculovirus system is phosphorylated by an unknown mechanism on the conserved tyrosine that is required for dimerisation and the DNA binding activity of the molecule. This indiscriminate phosphorylation suggests that the protein may also be inappropriately modified on other residues. The phosphorylation status of proteins including Stats is known to affect their migration through polyacrylamide gels (Shuai et al., 1992), and although rStat5 appears to migrate as a discrete band through a non-reducing NuPAGE protein gel (Figure 4.8), in a gel run

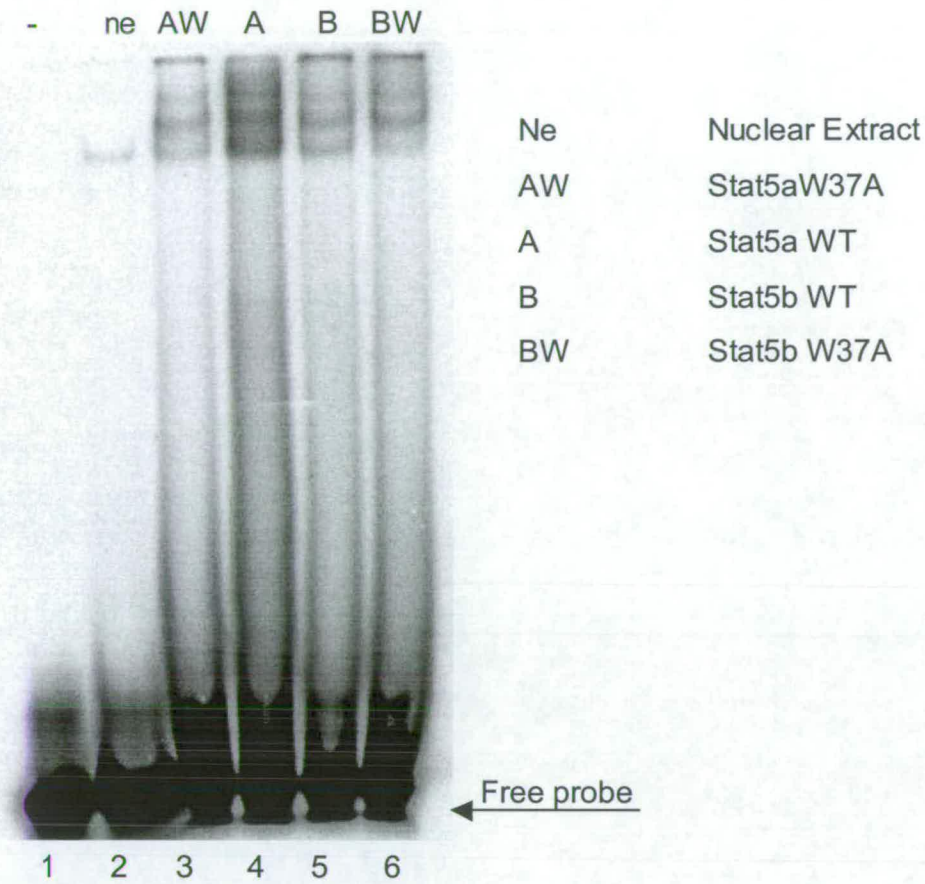


Figure 4.10 rStat5 Bandshifts of Oligonucleotide Probes.

Bandshifts of the oligonucleotide probe StM with recombinant Stat5aW37A, wild type rStat5a, wild type rStat5b, rStat5bW37A or with a mouse mammary nuclear extract. The first lane is the StM probe without any protein. An arrow at the bottom of the gel marks the migration of the free probe.

under reducing conditions Stat5 migrates as a collection of bands (Figure 4.9). Alternatively-phosphorylated forms of Stat5 may behave differently in a native polyacrylamide gel, such as those used for bandshifts. It is possible that alternative phosphorylation is the source of the multiple Stat5 bandshifts observed.

The rStat5 purified in Figure 4.8 migrates as a discrete band, but preparations that have been stored for long periods of time occasionally do show degradation products when separated on protein gels (data not shown). As the extra bands observed in bandshifts are seen to migrate more slowly than the primary Stat5 bandshift it is unlikely that degraded Stat5 binding is the source of the multiple bands. Two bands were also observed in bandshifts with nuclear extracts. These migrate at the same speed as the rStat5 bands although the relative intensity varies between individual extracts and rStat5 preparations (Figure 4.10 and Figure 4.2).

Recombinant Stat5 carrying the W37A mutation shows a similar bandshift pattern of the oligonucleotide probes as their wild type counterparts. This sugests that the multiple bands are not due to the previously characterised tetramerisation interactions. The possibility still exists that Stat5 interacts with other Stat5 molecules through an uncharacterised mechanism.

Previous tetramerisation interactions (John et al., 1999; Vinkemeier et al., 1996; Soldaini et al., 2000), required that binding sites for Stat5 be spaced between 6 and 10 bp apart in order to facilitate tetramerisation interactions. The three binding sites on nAB are spaced much farther apart than this with 68bp between A1 and A3 and 117bp between A3 and StM. Therefore it is

unlikely that tetramerisation interactions are involved in Stat5 binding to the naked probe nAB, although half Stat5 consensus sites (TTC or GAA) are present near to both the A1 and the A3 binding sites. These are spaced 11 and 12 bp from their counterpart in the defined binding site, this spacing is still compatible with the formation of higher order Stat5 structures. Also on certain sequences Stat5 preferentially forms tetramers rather than dimers (Soldaini et al., 2000).

As there are three binding sites present, up to three Stat5 dimers should be able to interact with nAB separately. The StM binding site is a strong site and will bind Stat5 preferentially compared to the weaker A1 or A3 binding sites. Because Stat5 bandshift products on even the short probe StM give rise to multiple bands, it is impossible to say with any certainty how many Stat5 binding sites are occupied solely by the migration through the gel.

4.5.2 Recombinant Stat5 Bandshifts on nAB.

Stat5 bandshifts on the nAB probe show a similar pattern to those on the shorter oligonucleotide probe StM (Figure 4.11), although they appear to migrate at a slightly slower rate. This is likely to be due to the extra retardation effects of the longer DNA sequence. There is a slight difference in migration between the two rStat5a species, wild type (A) and W37A (AW) and the two rStat5b species, B and BW. Stat5a migrates at a slightly slower rate than Stat5b, but Stat5a is also slightly larger than Stat5b (787 amino acids as opposed to 794 amino acids). The third slow migrating band seen previously with the oligonucleotide bandshifts is stronger in shifts with Stat5a than it is in shifts with rStat5b. As the Stat5 will eventually be used to shift nucleosome associated probes, two lanes from a gel run under the conditions nucleosome reconstitute gels are run under are shown in Figure

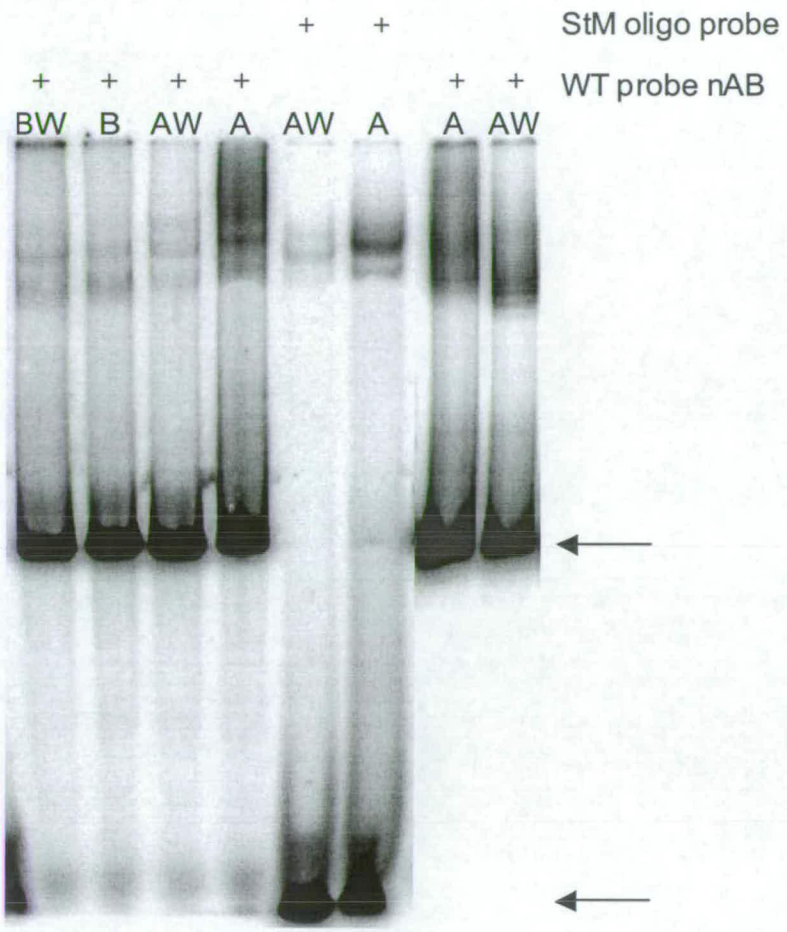


Figure 4.11 rStat5 Bandshifts of nAB and StM.

Bandshifts of the oligonucleotide probe StM and of probe nAB by rStat5 as indicated. The last two lanes are from a different gel, which was run under the conditions used for reconstitution gels rather than those of the other Stat5 bandshift gels to highlight the difference in migration seen between gels. Arrows to the right of the gel mark the migration of the free probes nAB (upper arrow) and StM (lower arrow).

4.11 (last two lanes) to compare migration behaviour. When bandshifts of the probe nAB are run under the same conditions as the reconstituted nucleosomes, a different migration pattern is observed (Figure 4.12 and 4.11). For the Stat5 bandshifts described above, 5′ end-labelled nAB was incubated with 1μg rStat5 and separated on a 6% 1 X TBE polyacrylamide gel run in 0.25 X TBE buffer at 200V constant for three hours at room temperature. Reconstituted nucleosomes are separated on a 5% 1 X TBE polyacrylamide gel in 1 X TBE buffer, at 60V constant overnight (16 hours) at +4°C. This DNA is labelled by addition of [α^{32}P] dCTP into the PCR reaction that produces nAB. Under these conditions the three Stat5 bandshift products do not resolve as well and the difference in migration between Stat5a and Stat5aW37A appears to be more pronounced (Figure 4.12). Any of the changes in the way the gels are run could influence the resolution.

4.6 Stat5 Binding Sites on the BLG Promoter.

Stat5 binding sites in nAB have been mutated by site directed mutagenesis both individually and in various combinations to the same sequences used in the mutate oligonucleotide probes (Figure 4.1). There are a total of eight different possible combinations of binding sites (Figure 4.13). Using these mutated nAB probes the migration of Stat5 bandshifts with varying numbers of Stat5 binding sites was studied. It was found that, similar to the histone core, the position of a bound Stat5 dimer on the DNA fragment affected the migration through the gel. In Figure 4.12 only one Stat5 dimer can bind to each of A1S (lane 2) and A3S (lane 3) but the A3S complex migrates faster than the A1S complex. On the A1S probe, Stat5 can only bind at the centrally positioned A3 binding site (Figure 4.13) while on the A3S probe, binding can only occur at the A1 position, which is situated towards the end of the fragment. It would appear that as with a nucleosome reconstitution, a

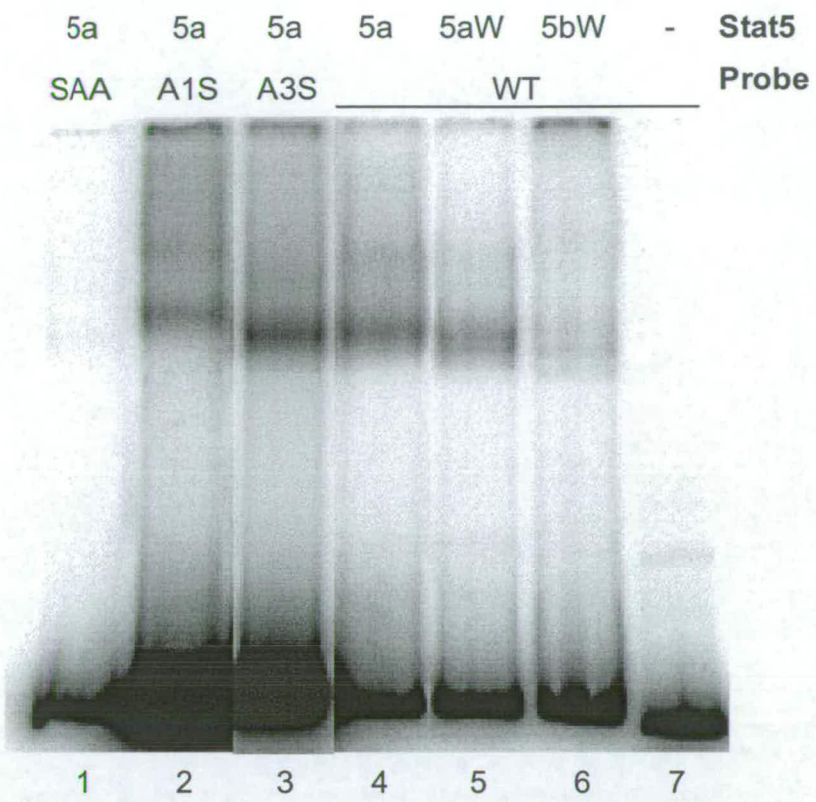


Figure 4.12 Bandshifts of nAB.

Bandshifts of variations of the probe nAB with rStat5 as indicated below were separated on a 5% polyacrylamide gel in 1 x TBE overnight at 60V constant, these are the conditions used for nucleosome bandshifts, but these shifts are of naked probes. Lanes two and three contain probe what was labelled with ^{32}P from a different activity date, so although the amount of DNA added to each reaction is equal, the specific activity is relatively greater in these two lanes. The intensity of signal in lane three has been adjusted to be able to compare band migration in this gel. Lanes 1-4 contain WT Stat5a, lane 5 contains Stat5a W37A, lane 6 Stat5bW37A and lane 7 no protein. Lane 1 is a shift of the SAA version of the fragment, Lane 2 is A1S, lane 3 is A3S and lanes 4-7 are WT.

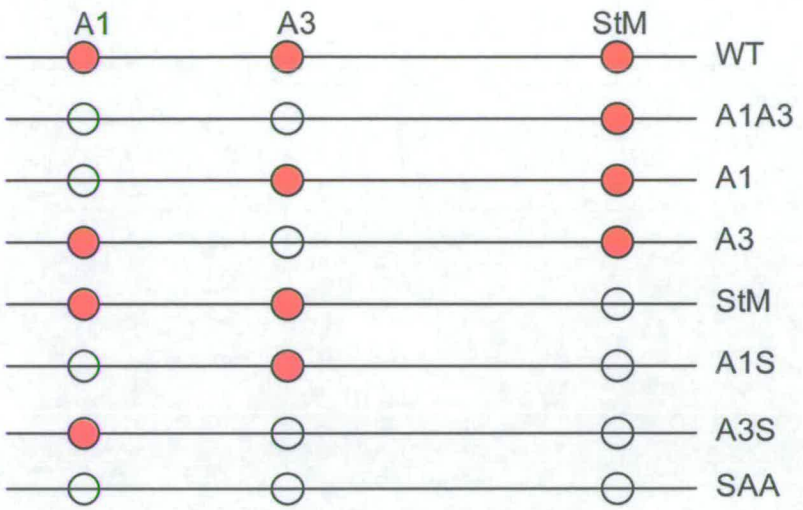


Figure 4.13 Combinations of Stat5 Binding Sites.

Representation of nAB showing the eight possible combinations of Stat5 binding sites. Red circles represent wild type binding sites for Stat5, while open circles represent mutated sites. The probes are named for the mutated binding sites in each.

centrally positioned Stat5 migrates through a native polyacrylamide gel at a slower rate than an end positioned Stat5. The effect does not appear to be as pronounced with Stat5 as with a nucleosome core. This may reflect the way the two protein complexes bind DNA, in particular the angle of DNA entry and exit. Stat dimers appear to simply bind with only a minor conformational change in the DNA, while in a nucleosome core the DNA is bent round the protein complex. (Becker et al., 1998; Chen et al., 1998; Richmond and Davey, 2003; Luger et al., 1997).

Unexpectedly we are unable to determine the number of Stat5 molecules bound to the probe by the presence of a faster migrating band in shifts of nAB containing only one intact binding site compared to the WT with three Stat5 biding sites (Figure 4.12, compare shifts of wild type (lane 4) and A3S nAB (lane 3)). This will also explain why there is little difference between binding of WT and W37A Stat5. nAB is not predicted to facilitate a tetramerisation interaction by the spacing of the Stat5 sites. My hypothesis is that as the StM site is a much stronger binding site, Stat5 binds preferentially at this site and the binding observed is just to this one site. The majority of the probe remains unbound. This free probe contains a large reservoir of StM binding sites, which Stat5 will always preferentially bind to before A1 or A3. In Figure 4.2, A1 and A3 did not completely compete Stat5 binding to StM at even 200X excess indicating that Stat5 affinity for StM is at least 200 fold greater than for either of the weaker sites. This theory was tested in a titration experiment where the amount of probe to rStat5added was reduced to a 1:1 ratio. However the Stat5 shifted probe band reduced in intensity in line with the free probe and no extra bands could be detected (data not shown).

The relative strength of the StM, A1 and A3 binding sites is also indicated in Figure 4.12. In this figure the SAA and WT probes are labelled using ^{32}P with a lower specific activity than the other fragments, lanes have been adjusted accordingly. The WT nAB in lane 4 contains all three Stat5 binding sites, and lanes 2 and 3 contain only one weak site. The total amount of signal present is much reduced in lane 4 compared to lane 2 (A1S), but the intensity of the shifted band is about equal indicating WT binds more Stat5 than A1S does. Similarly the total signal (after adjustment) in lane 3 (A3S) is greater than in lane 4, but the shifted band is only slightly more intense. These results could also be explained by the multiple binding sites present on nAB, but given the known greater affinity of Stat5 for StM it is likely to be due to StM binding only (Figure 4.2 and (Burdon et al., 1994b)). The total signal in lane 3 (A3S) is less than in lane 2, but the shifted band is more intense, suggesting that the A1 binding site is stronger than the A3, but not as strong as StM.

As expected, there is very little or no Stat5 binding to the SAA probe (Figure 4.12 lane 1), as this has all three Stat5 binding sites mutated. All three mutations have been shown individually not to bind Stat5 in Figure 4.2.

4.7 Summary of rStat5 Binding.

The above experiments show that recombinant Stat5 produced in the baculovirus expression system behaves in a similar manner to Stat5 from a nuclear extract in bandshift experiments. Under the correct conditions, sufficient levels of phosphorylation were found to use the recombinant protein in bandshift reactions without further action to phosphorylate Stat5. rStat5 binds to nAB producing a distinct reproducible bandshift. Bandshifts with the rStat5aW37A mutants and rStat5b appear to run faster, indicating a

smaller complex, than shifts with the WT rStat5a on the long probe. This may indicate a failure of the rStat5aW37A mutants and Stat5b to make Stat5-Stat5 (dimer-dimer) interactions. Stat5b may not have the same tendency as Stat5a to tetramerise (John et al., 1999). Recent reports suggest that Stat proteins can interact with each other without the requirement for activation (Schroeder et al., 2004; Ota et al., 2004). There are also implications that the W37A mutation is not vital in making tetramerisation interactions and any affect it has is simply due to a destabilisation of the N-terminus of the protein (Chen et al., 2002). This is discussed further in Chapter 6.

5 STAT5A BINDING TO A CHROMATIN TEMPLATE.

The binding of both Stat5 and the histone octamer to a fragment of the BLG promoter, nAB, has been characterised in Chapters 3 and 4. To investigate Stat5a binding to nucleosomes, isolated positioning isomers were bandshifted by Stat5a. Relatively few studies have studied the interactions of transcription factors with chromatin in this way; a brief review of some of those that have is included below (Section 5.1). The lack of such studies is surprising when you consider that all interactions of transcription factors, and the DNA replication and transcription complexes with eukaryotic DNA must occur within a chromatin context.

5.1 Transcription Factor Interactions with Chromatin.

The organisation of DNA into chromatin has an intrinsic negative effect on gene expression, as the association of DNA with nucleosomes must adversely affect the binding of transcription factors. Nucleosome positions have been mapped both *in vivo* and *in vitro* over the promoter regions of many genes. Well studied genes include the chicken β globin gene (Kefalas et al., 1988; Yenidunya et al., 1994), the mouse mammary tumour virus long terminal repeat (MMTV LTR) (Richard-Foy and Hager, 1987), the yeast PHO5 locus (Lohr, 1997) for review and Ovine BLG (Boa, 1999; Gencheva and Allan, 2005). The 5S RNA gene has been extensively studied as its repeats contain strong nucleosome positioning signals (Gottesfeld, 1987; Meersseman et al., 1991; Flaus et al., 1996). These maps can indicate which transcription factor binding sites are covered by nucleosomes and which are situated in linker DNA. For example nucleosome positions on the 5S RNA gene can affect the binding of TFIIIA (Rhodes, 1985; Panetta et al., 1998). As such these maps are useful sources of information regarding how positioned nucleosomes may affect gene regulation at this fundamental level.

In reality, the activation of genes is more complicated and involves a complex interaction of transcription factors with each other and the chromatin template. First, transcription factors may or may not be able to associate with nucleosomal DNA. Bound factors may recruit other factors resuling in the modification of both these and histones. Large multi-protein assemblies such as the chromatin remodelling complexes may be recruited, resulting in nucleosome movement or displacement. Changes in the chromatin structure over regulatory regions can be seen with the activation or silencing of many genes, manifested by the appearance of DNaseI hypersensitive sites. All these processes also must occur in the context of the 30nm fibre, as well as within other higher order chromatin structures (Horn and Peterson, 2002; Adkins et al., 2004).

Several studies have now looked further at the interactions of transcription factors with nucleosomal DNA. The nucleosome structure over the MMTV LTR has been studied in relation to binding of the glucocorticoid receptor (GR). Changes in the DNaseI digestion patterns through a positioned nucleosome on the MMTV LTR in response to GR were observed (Richard-Foy and Hager, 1987; Perlmann and Wrange, 1988). More recent studies include the binding of HNF3 and the estrogen receptor (ER) to the vitellogenin promoter. Using DNaseI digestion and bandshifts of a reconstituted chromatin probe it was shown that HNF3 binds to nucleosome associated DNA at the periphery of a nucleosome (Robyr et al., 2000). Again on the vitellogenin promoter, estrogen receptor α (ERα) was shown to be more efficient than ERβ in the estrogen-stimulated activation of the *Xenopus* vitellogenin A2 estrogen response element (ERE) on a chromatin template,

using an *in vitro* transcription system with a HeLa nuclear extract (Cheung et al., 2003).

The *Xenopus* 5S RNA gene contains strong nucleosome positioning sequences. The oocyte and somatic versions of the gene are differentially activated by the same factors, in a manner regulated by specific nucleosome positioning. TFIIIA will preferentially bind in the nucleosome positions found in the somatic gene, but only if the 3' end of its binding sequence is located outside the nucleosome. Nucleosomes on the oocyte gene position more over the TFIIIA binding sequence and preferentially bind the linker histone H1 (Rhodes, 1985; Panetta et al., 1998). This binding invoves displacement of ~20bp of DNA from the edges of the nucleosome (Vitolo et al., 2004).

NF1 binds to the MMTV LTR after hormone induction of the GR *in vivo,* and appears to do so on DNA associated with a nucleosome. However NF1 will not bind to nucleosomal DNA *in vitro,* but will bind to DNA associated with an H3/H4 tetramer, representing a nucleosome remodelling event. Both active GR and progesterone receptor will bind to nucleosomal DNA on the MMTV LTR and binding of these appears to be responsible for a nucleosome change that allows NF1 to bind. This may involve recruitment of the SWI/SNF complex (Piña et al., 1990; Cordingley et al., 1987; Eisfeld et al., 1997; Ostland Farrants et al., 1997; Spangenberg et al., 1998). These interactions may be relevant to this project as MMTV is a mammary specific promoter and there are binding sites for both NF1 and GR on the BLG promoter.

The activating potential of Stat1 within a chromatin template has been studied using an *in vitro* transcription system. This showed the C terminal domain to be essential for transcription in a chromatin context, and that this required the recruitment of the transactivator CBP/p300. The binding of Stat1 to the chromatin template, but not necessarily to nucleosome associated DNA, was demonstrated by DNaseI protection (Zakharova et al., 2003).

The majority of these studies only show that transcription factors can bind within chromatin, not how they interact with individual positioned nucleosomes. Authors are reluctant to commit as to whether or not their transcription factor binds to DNA in a nucleosome. In this chapter, the binding of Stat5a to sites incorporated in the previously characterised nucleosome positioning isomers on the BLG promoter is investigated, and a potential role of nucleosomes in potentiating higher order Stat5a-Stat5a interactions questioned.

5.2 Stat5a Binding to a Reconstituted Chromatin Template.

There are three sets of nucleosome positions on the fragment of the BLG promoter (nAB) that are available for analysis; characterisation of these was described in Chapter 3. All three positions cover the central Stat5 binding site, A3. The strongest two nucleosome position, N1 and N2 cover only A3 (Figure 5.1). The N2 position also borders the most downstream Stat5a binding site (StM) (Figure 5.1). The third nucleosome position, N3, is a combination of end-positioned nucleosomes and sequence directed nucleosome positions that sit over both the A3 site and the most upstream Stat5 binding site, A1 (Figure 3.16 and Figure 5.1). With this knowledge nucleosomes were reconstituted onto variations of nAB that contain combinations of the various Stat5 binding site mutations (Figure 4.13) to find

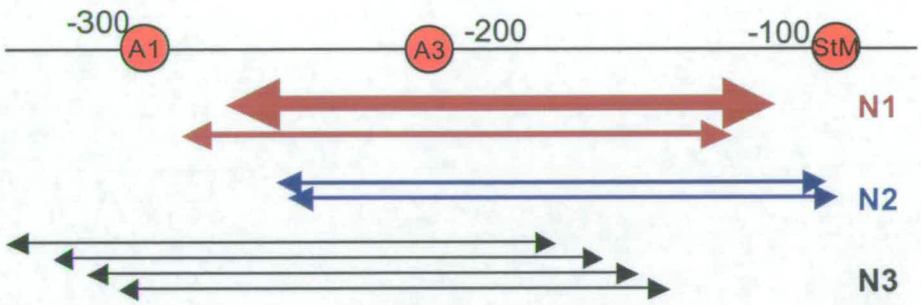


Figure 5.1 Relation of Nucleosome Position Isomers to Stat5 Binding Sites.

Positions of nucleosome position isomers on nAB, determined in Chapter 3, are represented by arrows. The positions are represented as groups that migrate as N1, N2 or N3. The relative strength of each position is represented by the thickness of the arrow. Stat5 binding sites A1, A3 and StM are represented by red circles. See Figure 3.16 for the precise nucleosome positions.

out whether or not Stat5a can bind to DNA associated with a nucleosome, and if so which Stat5a binding sites are available for binding on each nucleosome position.

5.2.1 Reconstitution onto nAB Probes.

Nucleosomes were reconstituted onto the WT, A1A3, SAA, A1S, A3S, and A3 versions of nAB (See Figure 4.13 for naming of nAB mutations), and the individual positioning isomers isolated (Figure 5.2). Some of the reconstitutes contain three bands (marked by solid arrows) as was observed previously for WT nAB (Figure 3.5). Others contain an extra band (NX marked by an open arrow in Figure 5.2) that migrates between the N2 and the N3 positions. The reconstitutes containing NX resemble the reconstitute in Figure 3.8, lane 5 and the NX band is also similar to that formed during ExoIII digestion of N2 (Figure 3.10 lane 7). The gel in Figure 5.2 was run shortly after the isolated positions were separated, there is no certainty that the isomers remain in these positions. Redistribution can and does occur after longer periods of storage, as observed in nucleosome isomer bands in Figures 5.5 through 5.10. However the majority of the signal remains in the correct position.

5.2.2 NX and the A3m Mutation.

The N2 band in the reconstitute on WT nAB accounts for 29.6% of the total signal. In the reconstitute on A1A3, which contains the extra band, the N2 band and the NX band together account for 29.5% of the total signal (17.3% and 12.2% respectively in N2 and NX). The similarity of the signal distribution between N2 and N2/NX suggests that NX is formed by a movement from the N2 nucleosome. The proportion of signal in the N1 and N3 bands remains unchanged.

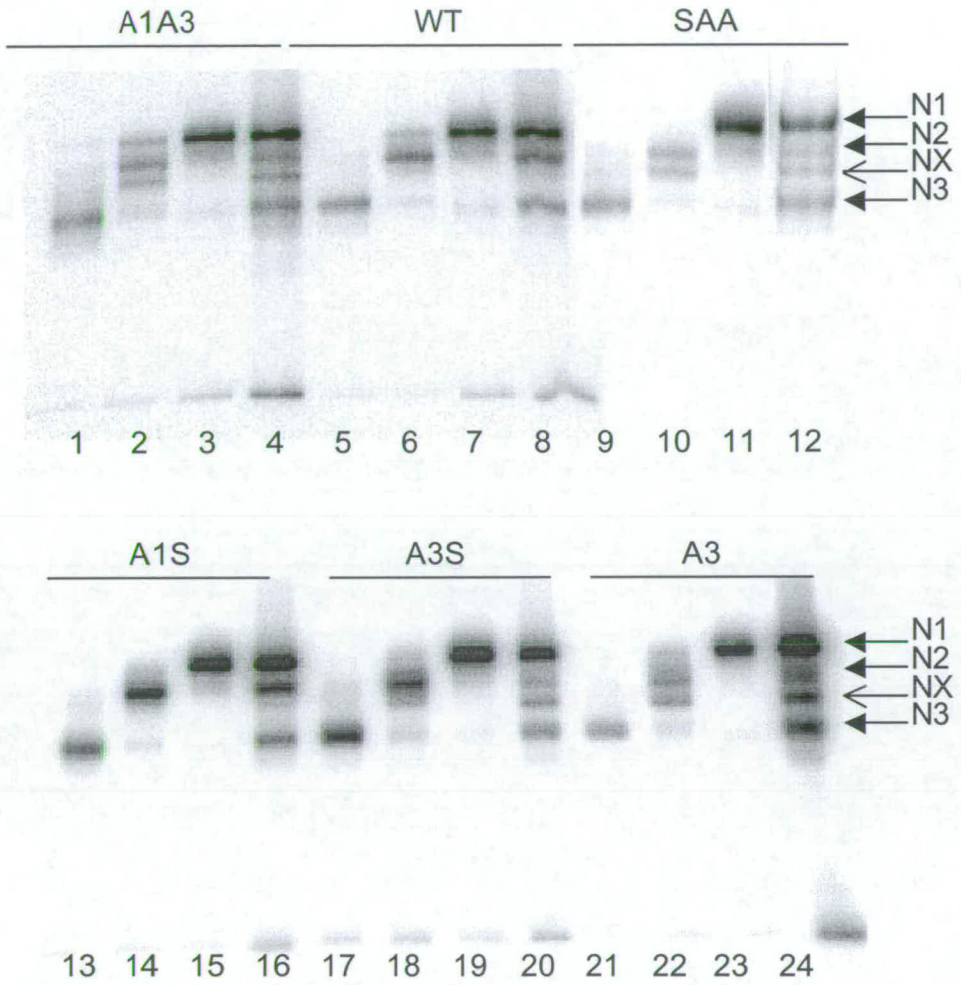


Figure 5.2. Nucleosome Positioning Isomers from Reconstitutions onto nAB.

Nucleosomes were reconstituted onto variations of the probe nAB containing various combinations of mutated or wild type Stat5 binding sites. Nucleosome positioning isomers were isolated for each (Lanes 1-4 are nucleosomes reconstituted onto A1A3; Lanes 5-8 on WT; Lanes 9-12 on SAA; Lanes 13-16 on A1S; Lanes 17-20 on A3S and Lanes 21-24 are on A3. Solid arrows at the right hand side mark the three main nucleosome positions N1, N2 and N3. An open arrow marks a fourth position present only in reconstitutes on probes that contain the A3 mutation. Bands were separated on 5% non-denaturing polyacrylamide gels run in TBE.

The NX band is present only in reconstitutes formed on an nAB probe that contains the mutated A3 Stat5a binding site (i.e. A3S, A3, SAA and A1A3) and is not seen in reconstitutions where A3 is wild type (i.e. WT and A1S). Thus it appears likely that the A3m mutation is causing a change in the nucleosome positions on nAB. The A1m and StMm mutations do not visibly affect the nucleosome positions. The migration of the nucleosome position isomers N1 and N3 appears unchanged by the A3m mutation.

Mutation of the A3 Stat5 binding site to A3m involved mutation of an A to a T residue at –204 bp from the transcription start site and also an insertion of a T at –206 (Figure 4.1). Of the three Stat5 binding site mutations designed, only the A3 mutation contains an insertion. The A1 and StM mutations contain only straightforward nucleotide exchanges. It is possible that it is this insertion that is causing the observed change in nucleosome migration. It has been suggested that the 20-30bp either side of the dyad are crucial to nucleosome positioning (Figure 5.3) and (FitzGerald and Simpson, 1985). The mutated A3m site is in this region for the N2 mapped nucleosomes although it is closer to the dyad axis in the N1 mapped nucleosome. It may be that the insertion of an extra base pair in the A3m mutation is sufficient to cross a threshold in the statistical likelihood that the NX position will exist. The migration pattern with the A3m mutation is similar to that of the (*Rsa*I) reconstitution (compare Figure 5.2 lanes 4, 12, 20 and 24 to Figure 3.8 lane 5) suggesting a tendency of this fragment to position a nucleosome at NX, but that this is masked by the stronger N2 binding site. This may provide an explanation why quite minor changes in each are sufficient to cause an entirely different nucleosome migration, as also discussed in Section (3.4.1).

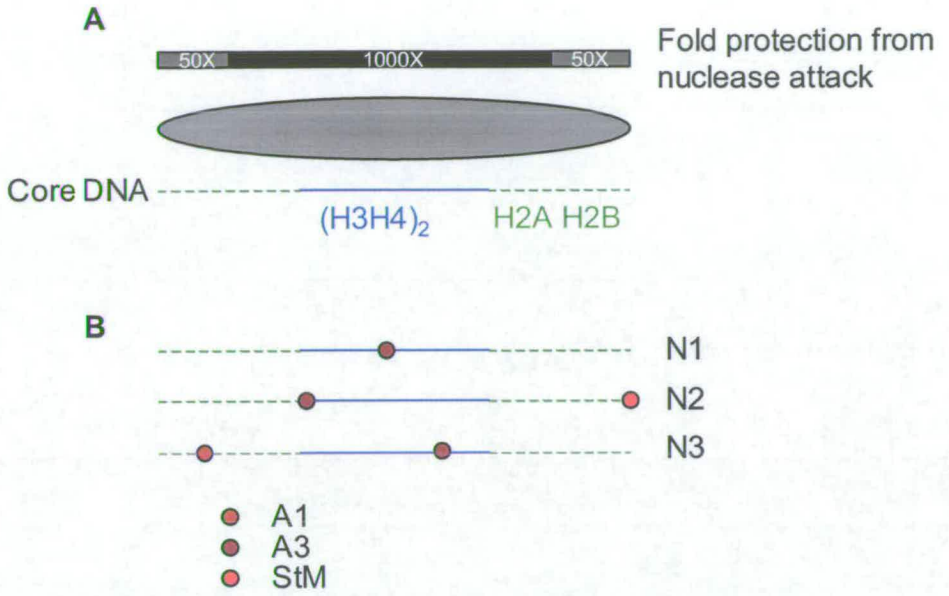


Figure 5.3 Selected Features of DNA Binding Within a Nucleosome.

A. The central 30 bp either side of the nucleosome dyad is bound by the $(H3H4)_2$ tetramer, this sequence is also responsible for directing nucleosome position. The outer DNA sequences are bound by H2A and H2B. The central 100bp is 1000 times less accessible than free DNA, whereas end sequences are 50 times less accessible (Linxweiler and Horz, 1984). **B**. Relation of Stat5 binding sites on core DNA to the H3H4 bound DNA, the three Stat5 binding sites are represented by different coloured circles.

A similar situation arises in a comparison of the sequences of the caprine and ovine BLG promoters. These contain subtle differences, which result in an altered nucleosome structure. The main *in vitro* mapped nucleosome in caprine is shifted by 10bp downstream placing its dyad at –174bp with respect to the transcription start site. The two promoter regions are very well conserved, but there is a substitution of two C residues in the ovine sequence to two T residues in the caprine sequence at 19 and 9 bp from the dyads respectively. There are nine other A/T to C/G mutations within the nAB region and one instance of a G residue in the ovine sequence that is not present in the caprine sequence (see appendices).

5.3 Stat5a Binding to a Whole Reconstitution.

Initial Stat5a bandshifts were carried out on unfractionated reconstitutes to demonstrate that Stat5a can bind to a reconstituted chromatin probe. Both WT Stat5a and Stat5a W37A bound to a reconstituted chromatin probe, as manifested by a broad band migrating more slowly than the isolated nucleosome positioning isomers (Figure 5.4A, lanes 2, 3, 5 and 6).

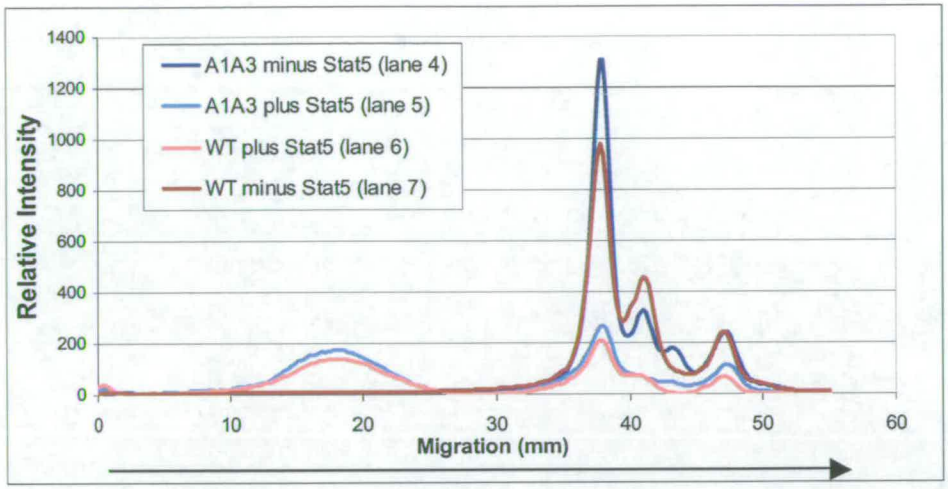
5.3.1 Migration of Stat5a Shifted Complexes.

A percentage of the shifted nucleosome band observed will be due Stat5a binding to naked nAB, as the nucleosome associated DNA was not separated from the free DNA. However the Stat5a shifted reconstitute band migrates more slowly than the Stat5a shifted naked DNA band does (Figure 5.4A compare lanes 1 and 2). This indicates a complex is formed of DNA bound to at least one Stat5a dimer plus a nucleosome, as opposed to Stat5a binding to naked DNA after the displacement of a nucleosome. Stat5a probably binds to the reconstituted fragment as well as to the naked DNA fragment judging by the relative disappearance of the naked DNA and nucleosome

Figure 5.4 Stat5 Bandshifts of Reconstituted Probes.

A. Reconstituted nAB (lanes 2-7), and naked DNA (lane 1) were bandshifted by the addition of recombinant Stat5a and separated on 5% non-denaturing polyacrylamide gels. Shifts of the WT and A1A3 nAB probes by both WTStat5a and Stat5aW37A are indicated above the gel. Populations of various shifted bands are marked by brackets either side of the gel. **B.** Signal trace following the migration of bands in lanes 4-7 from the gel in **A** after normalisation to lane loading (produced using AIDA software). The average signal intensity in each lane was plotted against the distance migrated in mm. The chart in **C.** represents the proportion of the signal that remains in each nucleosome position after Stat5 binding (lanes 5 and 6) expressed as a percentage of the signal in that nucleosome position before the addition of Stat5 (lanes 4 and 7) determined using Quantity One analysis of a phosphorimage. This indicates the extent each nucleosome positioning isomers is shifted by the addition of Stat5.

B.



C.

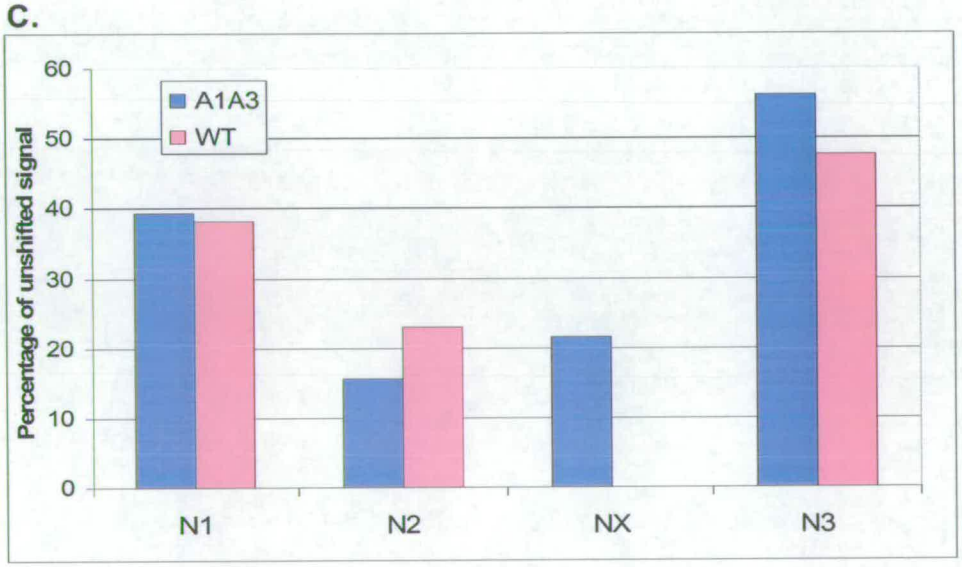


Figure 5.4 continued.

bands. This is not unexpected as all three nucleosome positions place the strongest Stat5 binding site (StM) external to the core DNA.

The shifted nucleosome band is very diffuse. As with the naked DNA shifts, no difference could be detected between the complex produced from shifts of either the WT or the A1A3 reconstitutes (Figure 5.4 lanes 5 and 6). Also there was no difference observed between the migration of complexes formed by the binding of WT Stat5a or Stat5aW37A. This is not unexpected as the shifted band is a complex structure, and little is known of the reasons why nucleosome position isomers migrate according to their position on DNA in non-denaturing polyacrylamide gels (Pennings, 1997). Also, the migration of Stat5a bandshifts of naked DNA are themselves not fully understood, involving also a certain amount of DA bending. Thus, the rate of migration of a complex formed by binding of Stat5 to a positioned nucleosome cannot be predicted and in any case will not be simple. The possibility also exists that Stat5 binding may result in repositioning of a nucleosome.

The graph in Figure 5.4B from left to right follows the trace of the signal intensity from top to bottom of lanes 4 through 7 of Figure 5.4A. In order to calculate which nucleosome positions Stat5a is binding to, the proportion of signal in each nucleosome position isomer was calculated as a percentage of the total signal in the lane. This was calculated for the reconstitutes on A1A3 and on WT both with and without addition of Stat5. The graph in Figure 5.4C shows the proportion of the signal in each nucleosome position in the Stat5 bound lanes, expressed as a percentage of the signal in that nucleosome position in the non-Stat5 lanes. These data indicate that Stat5 binds to all three nucleosome positions.

The proportion of signal remaining after the addition of Stat5a is lowest in the N2 and the NX positions, indicating that these nucleosomes bind Stat5a well. The proportion of the signal remaining is greatest in the N3 position and intermediate in N1. The conclusion drawn from this data was that Stat5a preferentially binds to N2/NX, more weakly to N1 and lastly to N3. This can also be seen by eye (Figure 5.4A). No major differences were detected between binding to the A1A3 probe or the WT probe. This is likely to be because the strongest Stat5a binding site, StM, is external to all of the positioned nucleosomes (Figure 5.1). Binding to the StM site probably accounts for the majority of the Stat5a binding detected.

The only factor influencing the binding of Stat5a to each positioned nucleosome on the A1A3 probe, which contains only the StM Stat5 binding site, is the proximity of this binding site to the nucleosome boundary. The order of Stat5a binding affinity is N2>N1>N3, likewise the proximity of StM to the nucleosome boundary is N2<N1<N3 with StM placed closest to the nucleosome in N2 and furthest from it in N3 (Figure 5.1). To investigate further the binding of Stat5a to each nucleosome position, Stat5a bandshifts were carried out on isolated positioning isomers.

5.4 Stat5a Binding to Isolated Nucleosome Positions.

Nucleosome positioning isomers isolated from reconstitutes on the variations of nAB (Figure 5.2) were used as probes in Stat5a bandshifts. Stat5a is shown to bind to all three nucleosome position isomers (Figure 5.5) on both the A1A3 and WT versions of nAB. A more detailed study of binding is carried out in Figures 5.6, 5.7 and 5.8. One limitation of the experiments described below is that, unless otherwise stated, each was conducted only once.

Nevertheless the results were consistent between several experiments lending weight to the conclusions drawn.

5.4.1 Nucleosome Redistribution.

It is also apparent that the nucleosome positions on the unbound reconstituted probe do not remain static, and that although the majority of the signal remains in the correct band, nucleosomes do redistribute to all three (or four in the case of A1A3) positions. For example, compare the migration of isolated nucleosome positioning isomers in Figure 5.2 with Figure 5.5. N3 is the most likely to remain at the correct position; this also is the case for untreated nucleosome positioning isomers (Section 3.2.4) and may reflect a tendency of nucleosomes to move towards an end position.

The redistribution observed is not due to the addition of Stat5a as the two outer lanes (1 and 14) in Figure 5.5 do not contain Stat5a, but yet display the same pattern of redistribution. Nucleosome redistribution is most likely due to spontaneous nucleosome movement brought about by the conditions the nucleosomes were stored under.

The percentage of the signal in each nucleosome position was calculated from a phosphorimage using Quantity One software for the two outer lanes containing the non-Stat5-shifted N1 or N2 positions (Figure 5.5 lanes 1 and 14). The total signal in all the nucleosome bands in each lane was taken as 100 and the proportion of signal in a defined area around each position calculated as a percentage of this. However because of the irregular shape of some bands, resulting in band overlapping, it may have been better to calculate the signal using lane scan integration. In the N1 isomer without Stat5a (Figure 5.5 lane 14) 77.3% of the signal remains in the N1 position and

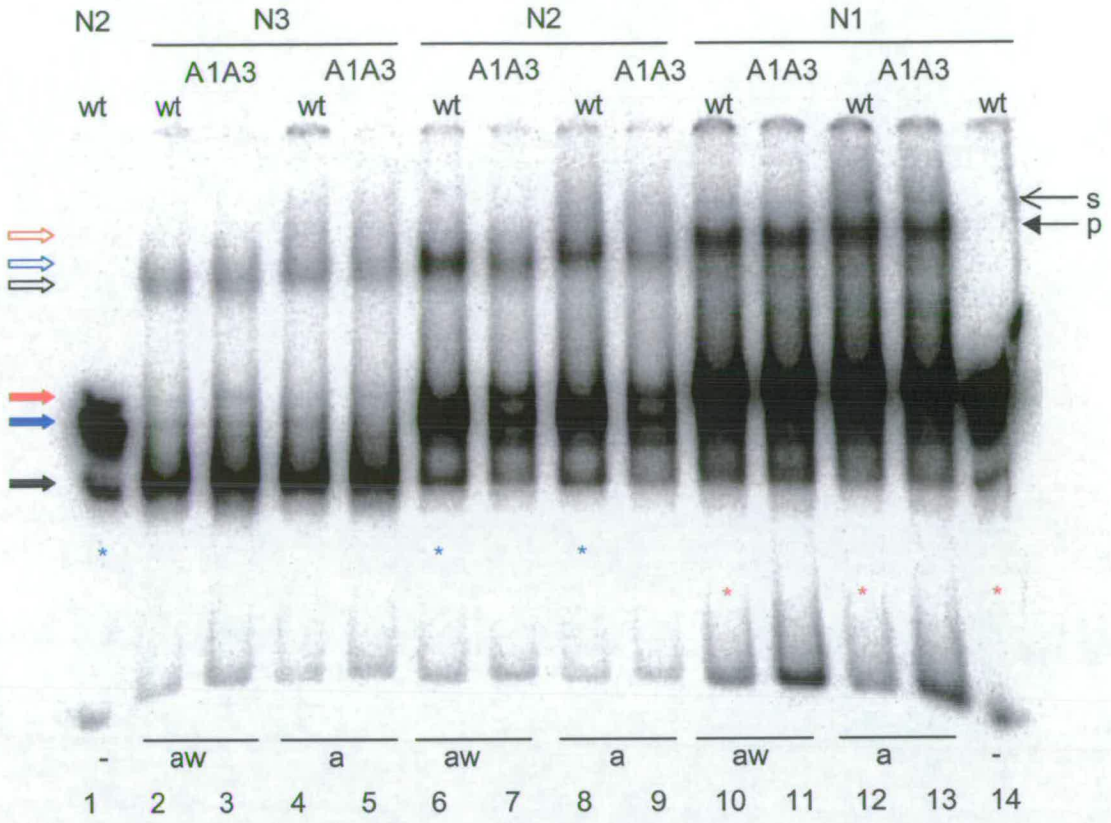


Figure 5.5. Stat5 Binding to Isolated Nucleosome Positions.

Isolated nucleosome positioning isomers N1, N2 and N3 were shifted by addition of either WTStat5a or Stat5aW37A and separated on non-denaturing polyacrylamide gels. Shifted probes are indicated at the top of the gel, and the Stat5a species added are indicated at the bottom. Block arrows on the left mark the migration of N1 (red) N2 (blue) and N3 (black) for shifted (open arrows) and non shifted (filled arrows) nucleosome positioning isomers. Stick arrows to the right of the gel mark the migration of the secondary (open arrowhead) and primary (closed arrowhead) bandshifts for nucleosome position N1. The percentage of signal in each nucleosome positioning isomer for the isolated positions N1* and N2* in the Stat5a shifted and non-shifted lanes was calculated by quantifying the signal in a set outline around each band using Quantity One software, results are shown in Table 5.1. The strengths of the Stat5a shifted bands were not quantified.

	N1	N2	N3
WT N1 – Stat5a	**77.3%**	18.5%	4.2%
WT N1 + Stat5a	**75.4%**	18.1%	6.5%
WT N2 – Stat5a	16.6%	**70.9**	12.5%
WT N2 + Stat5a	31.1%	**55.2**	13.7%

Table 5.1. Distribution of Signal in Isolated Nucleosome Positions .

The distribution of the signal calculated from a phosphor image using Quantity One Software in each nucleosome position for each isolated position in figure 5.5 is indicated, described as a percentage of the total signal in the nucleosome-shifted bands. This shows the extent of redistribution that has occurred between the positions. The 'correct' nucleosome position for each isolate is highlighted in bold.

in the N2 lane without Stat5a (Figure 5.5 lane 1) 70.9% of the signal remains in the N2 position (Table 5.1). The same analysis was carried out on the nucleosome distribution in the Stat5a shifted lanes. The proportion of signal in the N1 position in the N1 isomer lane after addition of Stat5a (Figure 5.5 Lanes 10 and 12) fell to 75.4% (from 77.3% in lane 14). The percentage of N2 contamination in the N1 isomer lanes remained constant on Stat5a addition (Table 5.1), suggesting the majority of the observed shift is of the N1 position.

In the N2 isomer lanes the percentage of the signal in N2 fell to 55.2% (from 70.9% in lane 1 without Stat5a) on addition of Stat5a (Figure 5.5 lanes 6 and 8). The percentage of the total signal in N3 for both the N1 and N2 isomers remained fairly constant regardless of Stat5a addition. The complete data is shown in Table 5.1. These data suggest that in the N2 bandshifts it is the N2 isomer that is shifted by Stat5a, but in the N1 shifts both N1 and N2 are shifted by binding of Stat5a. This would suggest that Stat5a binds to N2 better than it does to N1. The data from the two WT probes and the two A1A3 probes agreed, although only the data from WT lanes are shown in this analysis as the two outside lanes are on WT.

The migration of each Stat5a shifted nucleosome position is related to the migration of the isolated nucleosome position that was used as the probe. In Figure 5.5 the N1 Stat5a shift (open red arrow to the left of the gel) migrates more slowly than the N2 Stat5a shift (open blue arrow), which in turn migrates more slowly than the N3 Stat5a shift (open black arrow); just as their respective nucleosome positioning isomers do (filled red, blue, or black arrows). This suggests that the Stat5a shift observed is due to binding to the designated probe. The three Stat5a shifted nucleosome positions are not

separated to the same extent as the individual nucleosome positioning isomers are. This may be due simply to their increased size, the resolving ability of the gel, or to the position of bound Stat5a. The main difference is that the N3 position migrates much closer to the N1 and N2 positions in the Stat5a shifted nucleosomes than it does in the non-Stat5-shifted nucleosomes. The N3 nucleosome position is situated at the upstream end of the fragment, and the StM binding site at the downstream end. Stat5 binding to StM may even out the differences in the migration caused by the conformation of the DNA tails exiting the nucleosome resulting in a faster relative mobility of N3. The different migration of the three Stat5-shifted nucleosome positioning isomers accounts for the diffuse migration of the Stat5a shifted unfractionated reconstitutes in Figure 5.2, which are a combination of shifts from all the nucleosome positions. Much tighter bands are observed for the shifts of the isolated positioning isomers.

The migration of nucleosomes through non-denaturing polyacrylamide gels is governed by a combination of the mass and charge of the complex, but also by conformational factors. The free DNA tails exiting the nucleosome cause nucleosomes to migrate differentially according to their position on the fragment. Consequently a centrally positioned nucleosome will migrate at a slower rate than an end positioned nucleosome. Addition of glycerol to the gel mix removes this conformational factor, and nucleosomes migrate solely according to their mass and charge (Pennings, 1997; Pennings et al., 1992).

Conceivably, the migration of the Stat5 shifted nucleosome bands could be simplified by the addition of glycerol to the gel used to separate the shifted complexes. This would remove the conformation effect of the nucleosome

position, and each nucleosome bound to nAB would migrate at the same rate regardless of its position. This could be useful in further studies using bandshifts of nucleosomes and may make it possible to detect multiple Stat5a dimers bound to the probe. However, as the position of the nucleosome on the DNA is central to these studies of Stat5a binding, bandshifts were carried out on non-denaturing polyacrylamide gels that did not contain glycerol to be able to verify that nucleosomes remain positioned at the expected sites as indicated by their migration through the gel.

5.4.2 Stat5a Binding to Nucleosome Positioning Isomers on WT and A1A3 Probes.

A comparison of the binding of WTStat5a and Stat5a W37A to each isolated nucleosome position on both wild type and A1A3 nAB was carried out to look for binding of multiple Stat5a dimers (Figure 5.5). The shifts with WT Stat5a have a weak secondary band above the primary bandshift, which is not apparent in the shifts with Stat5aW37A. This band also appears in shifts of the A1A3 nucleosome, although A1A3 contains only one binding site for Stat5a. It is likely that this secondary band is, as with shifts of naked nAB, not due to DNA binding induced tetramerisation but simply to other undescribed Stat5 interactions. On the naked DNA probe the Stat5aW37A mutant binds equally as well as the WT Stat5a does as regards both the primary and the secondary band, but in the nucleosome shifts Stat5W37A does not appear to produce the secondary band. The secondary band appears stronger in N1 shifts than for any of the other nucleosome positions, but N1 primary bandshifts are also much stronger than any other, with the exception perhaps of the WT N2 Stat5a shifts, so the intensity of the secondary band may simply be a reflection of the signal available. Comparison of the signal in the secondary and primary Stat5a shifted bands

was complicated by the fact that the secondary bands are little above the background signal from the primary shift, which is extremely variable between lanes.

5.4.3 Analysis of Primary and Secondary Stat5a Bandshifts of WT and W37A Stat5a.

To investigate further the possibility of multiple Stat5a dimers binding to the WT and A1A3 probes, bandshifts of each isolated nucleosome position by WT Stat5a and Stat5a W37A were carried out on gels with better resolution. These samples were loaded on 5% polyacrylamide gels in 1 X TBE as previous, but the wells were wider, and the running buffer was diluted to 0.25 X TBE. This change in running conditions resulted in a tighter migration of the shifted bands. These results show that, as was observed in Figure 5.5 for each isolated nucleosome position, N1 (Figure 5.6), N2 (Figure 5.7) and N3 (Figure 5.8) WT Stat5a produces a secondary band that is not obvious in Stat5a W37A shifts of the same probe. Stat5a W37A shifts also appear to migrate at a marginally more rapid rate than WT Stat5a shifts. A Stat5a shift of nucleosome positions N1 and N2 on the triple Stat5 mutation of nAB, SAA, is included as a control to demonstrate that the shifts are due to Stat5a interactions with the probe and not with the nucleosomes. No shift is observed of either of these SAA probes or of a whole reconstitution onto SAA (data not shown).

The proportion of signal in each band, primary or secondary, expressed as a percentage of the total shifted bands is shown below each shift lane in Figures 5.6, 5.7 and 5.8. For each nucleosome position the proportion of signal in the primary and secondary band is the same regardless of the DNA

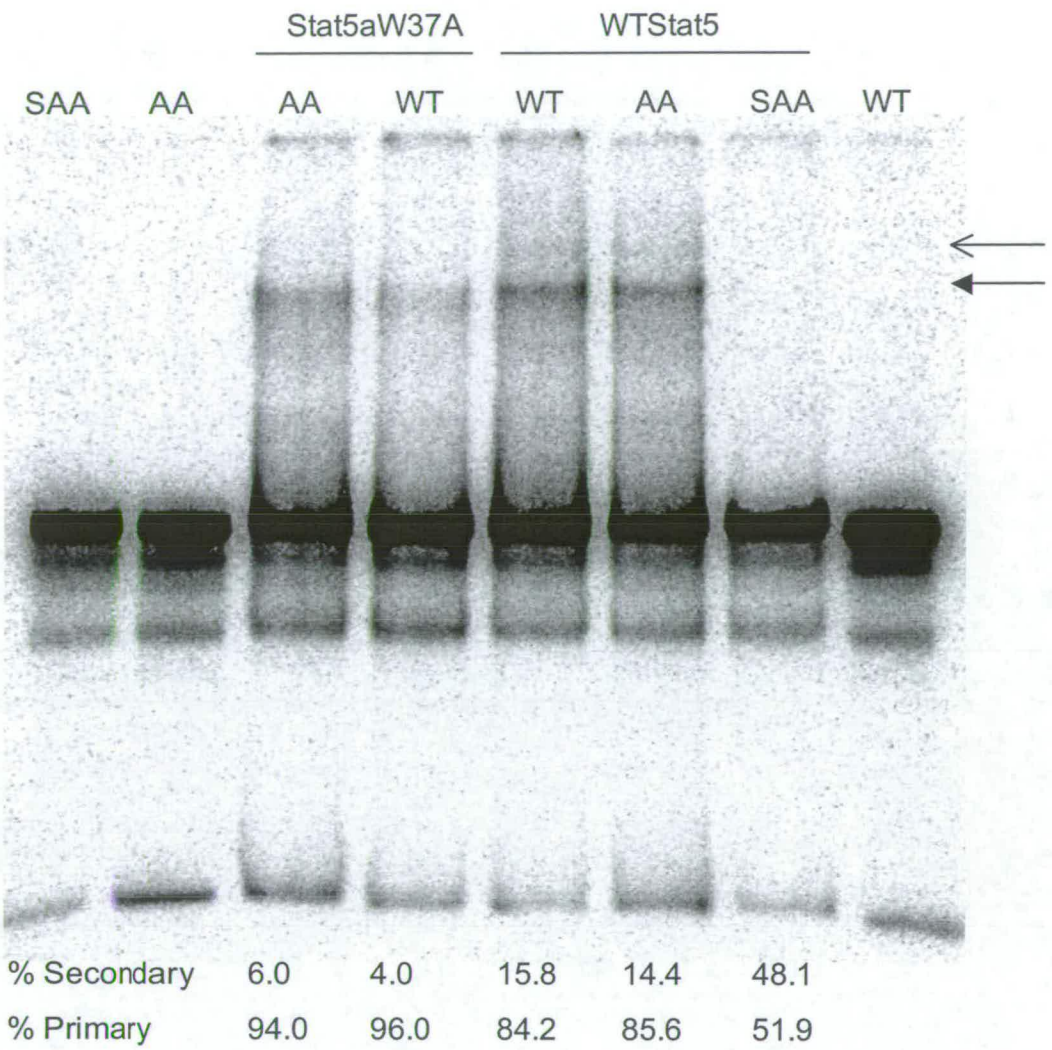


Figure 5.6 Bandshifts of Nucleosome Position N1.

The isolated nucleosome position N1 from reconstitutes onto either the WT, A1A3 (AA) or SAA versions of nAB was bandshifted by addition of 1μg of either WT Stat5a or Stat5a W37A. Complexes were separated on a 5% non-denaturing polyacrylamide gel in 0.25 X TBE at 60V constant overnight at 4°C. Analysis of the proportion of signal in the primary and secondary bands, expressed as a percentage of the total shifted band is shown underneath each shifted lane, including SAA to demonstrate that the weaker the shift the closer each value is to 50%. The signal is very weak so this quantification may be unreliable The Stat5a that was used to shift each nucleosome is indicated above the gel and arrows to the right of the gel mark the secondary (open arrow) and the primary (closed arrow) bandshifts.

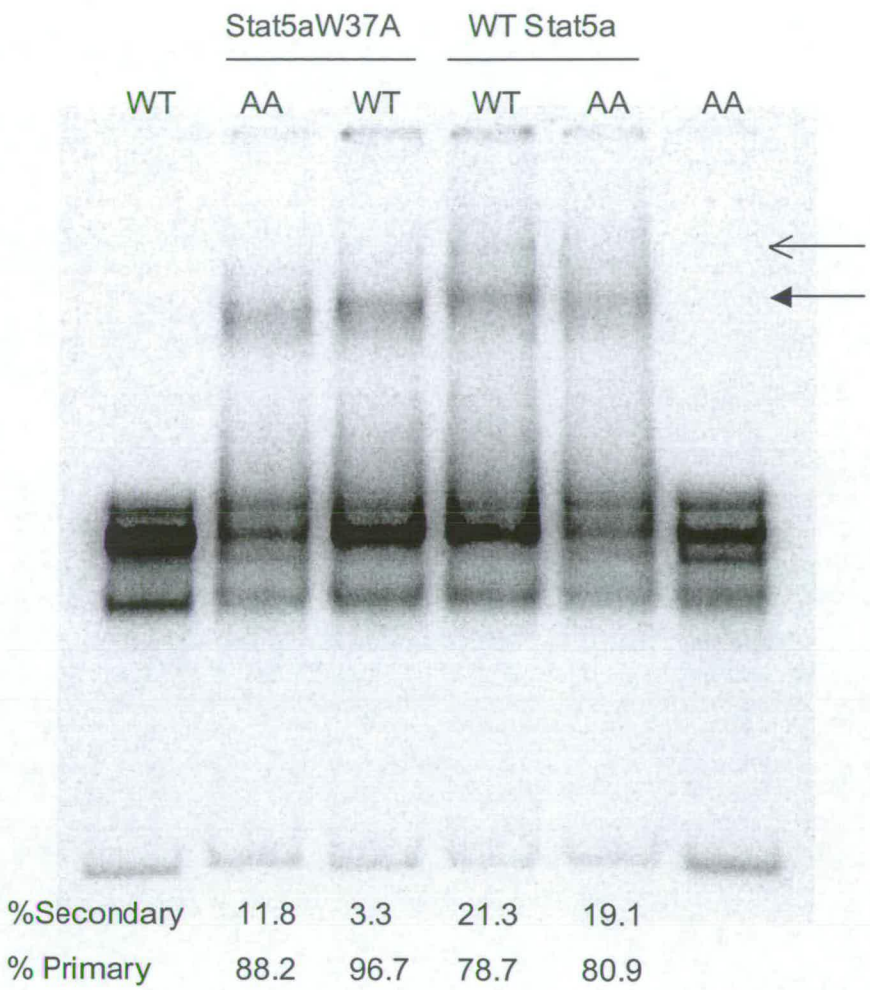


Figure 5.7 Bandshifts of Nucleosome Position N2.

The isolated nucleosome position N2 from reconstitutes onto either WT or A1A3 (AA) versions of nAB was bandshifted by addition of 1μg of either WT Stat5a or Stat5aW37A, and complexes were separated on a 5% non-denaturing polyacrylamide gel in 0.25 X TBE at 60V constant overnight at 4°C. Analysis of the proportion of signal in the primary and secondary bands, expressed as a percentage of the total shifted band is shown underneath each shifted lane. The signal is very weak so this signal may be unreliable. The Stat5a that was used to shift each nucleosome is indicated above the gel and arrows to the right of the gel mark the secondary (open arrow) and the primary (closed arrow) bandshifts.

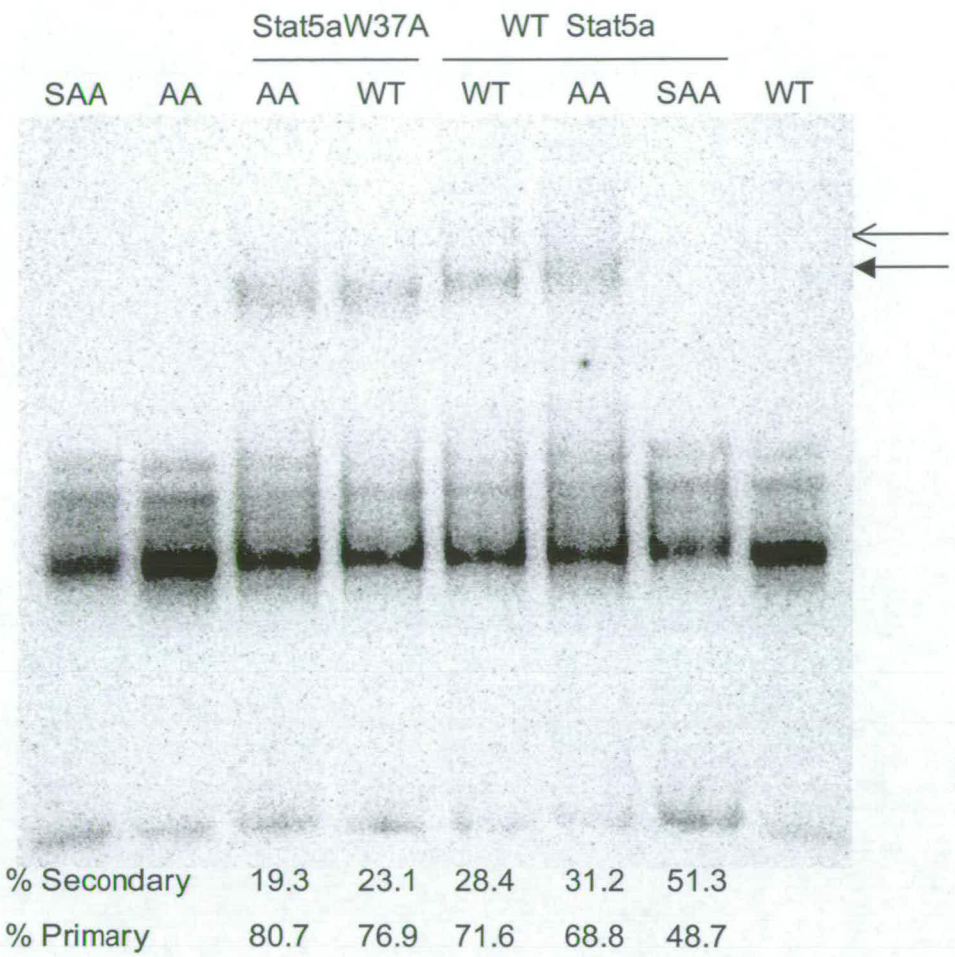


Figure 5.8 Bandshifts of Nucleosome Position N3.

The isolated nucleosome position N3 from reconstitutes onto either WT, A1A3 (AA) or SAA versions of nAB was bandshifted by addition of 1μg of either WT Stat5a or Stat5aW37A. Complexes were separated on a 5% non-denaturing polyacrylamide gel in 0.25 X TBE at 60V constant overnight at 4°C. Analysis of the proportion of signal in the primary and secondary bands, expressed as a percentage of the total shifted band is shown underneath each shifted lane, including SAA. The signal was very weak so this quantification is unreliable. The Stat5a that was used to shift each nucleosome is indicated above the gel and arrows to the right of the gel mark the secondary (open arrow) and the primary (closed arrow) bandshifts.

For example both A1A3 and WT nAB in the N1 nucleosome position contain 15% of the shifted counts in the secondary band when shifted by WT Stat5a, but only 4-6% when shifted by Stat5aW37A (Figure 5.6). This is approximately 3 times more in the secondary band in shifts by WT Stat5a than in shifts by Stat5aW37A. The N2 position has a similar ratio with 20% of the counts shifted by WT Stat5a migrating in the secondary band and 3-11% with Stat5aW37A shifts (Figure 5.7). This is 2.7 times more in WT compared to W37A Stat5a shifts. The N3 probe also follows the same pattern with about 30% in the secondary band in WT Stat5a shifts and about 21% in Stat5aW37A shifts (or 1.4 times difference between WT and W37A). The higher values in the N3 shift are likely to be caused by the fact that the N3 probe contains by far the lowest signal, and the signal in the shifted bands is little above background. As demonstrated by quantification of shifted bands in the SAA lanes in Figures 5.6 and 5.8, there is no Stat5a shifted signal and the signal is at background level, so the ratio becomes close to 50:50. The Stat5 shifted signal is weak in all three figures so the above quantification is unreliable, but nevertheless gives an indication of the signal distribution.

These experiments demonstrate that there is a link between the appearance of the secondary shifted band and the use of either WT Stat5a or Stat5aW37A, and that this is approximately 2 times more pronounced in the N1 and N2 nucleosome positions than the N3 position (WT/W37A ratios of 3 and 2.7 compared to 1.4 in N1, N2 and N3 respectively).

However these data are from only one set of gels, and although they agree with each other, the results may not be representative. The bandshifts in Figures 5.5, 5.6, 5.7 and 5.8 all show the secondary band is present in WT

Stat5a shifts but not in Stat5aW37A shifts, lending weight to the conclusion that there may be a difference between WT Stat5a and Stat5a W37A. More work must be carried out to determine if the increased appearance of the secondary band in shifts of the N3 position compared to the N1 and the N2 positions is real, and also to provide further insight into the observed difference between WT and Stat5aW37A.

These experiments have also demonstrated that Stat5a can bind to a reconstituted nucleosome for all three nucleosome positions on the WT and A1A3 probes. All three nucleosome positions place the strongest binding site, StM, external to the nucleosome and this site is present in both the WT and the A1A3 probes. To investigate binding of Stat5a to other sites in a nucleosomal context, mutations of nAB which abolish the StM and other Stat5 binding sites individually were employed (Figures 5.9 and 5.10).

5.5 Stat5a Binding to Sites Covered by a Nucleosome.

There is currently no evidence in the literature to suggest whether or not Stat5a, or any Stat family protein can bind to sites situated within a nucleosome. Stat1 will bind to chromatin, but the question remains open as to whether or not this occurs within DNA that is associated with a nucleosome (Zakharova et al., 2003).

5.5.1 Stat5a will not Bind to A3 Within a Nucleosome.

The A1S version of nAB contains only the central (A3) Stat5a binding site, StM and A1 having been mutated (Figure 4.13). The A3 binding site is covered by the nucleosome in every nucleosome position mapped on the nAB fragment (Figures 5.1, 5.3 and 3.17). When a nucleosome from a reconstitute on A1S is bandshifted by addition of WT Stat5a, nucleosome

positions N1 and N2 are not shifted (Figure 5.9 lanes 2 and 5). The same is true for nucleosome position N3, as an unfractionated reconstitute onto A1S also does not bind Stat5a (Figure 5.10 lane 2). As Stat5a will bind to naked A1S (Figure 4.12 lane 2) it must be the presence of the nucleosome that precludes binding of Stat5a to A1S. This data indicates that Stat5a does not bind to the A3 binding site when it is located within a nucleosome. The minimum distance of the A3 binding site from a nucleosome boundary is 40bp, this is calculated from the upstream boundary of the N2 position at –169bp from the transcription start site. The A3 Stat5 binding site covers the sequence from –211 to –201 and the upstream boundary of the nucleosome is at –242bp from the transcription start site. A binding site 40bp from the nucleosome boundary is located fairly centrally within the nucleosome. It is possible that a binding site located closer to the nucleosome boundary may allow Stat5a to bind.

5.5.2 Binding of Stat5a to A1 Within Nucleosome Associated DNA.

To investigate Stat5 binding to sites closer to the nucleosome boundary, similar experiments were carried out for the A1 Stat5a binding site as for the A3 site, using the A3S version of nAB. This has both the A3 and the StM Stat5 binding sites mutated (Figure 4.13). The A1 binding site is located outside the nucleosome in positions N1 and N2, but internal to the nucleosome in position N3 (Figures 3.17 and 5.1), although the location of Stat5 binding sites in the alternative band NX are unknown. Unexpectedly, given that previously Stat5a has been shown not to bind to DNA associated with a nucleosome (Section 5.5.1), Stat5a will bind to an A3S reconstituted probe for all three nucleosome positions including N3 (Figure 5.10). This suggests that the position of a Stat5 binding site with respect to the nucleosome boundary and dyad are important in Stat5a binding. The A1

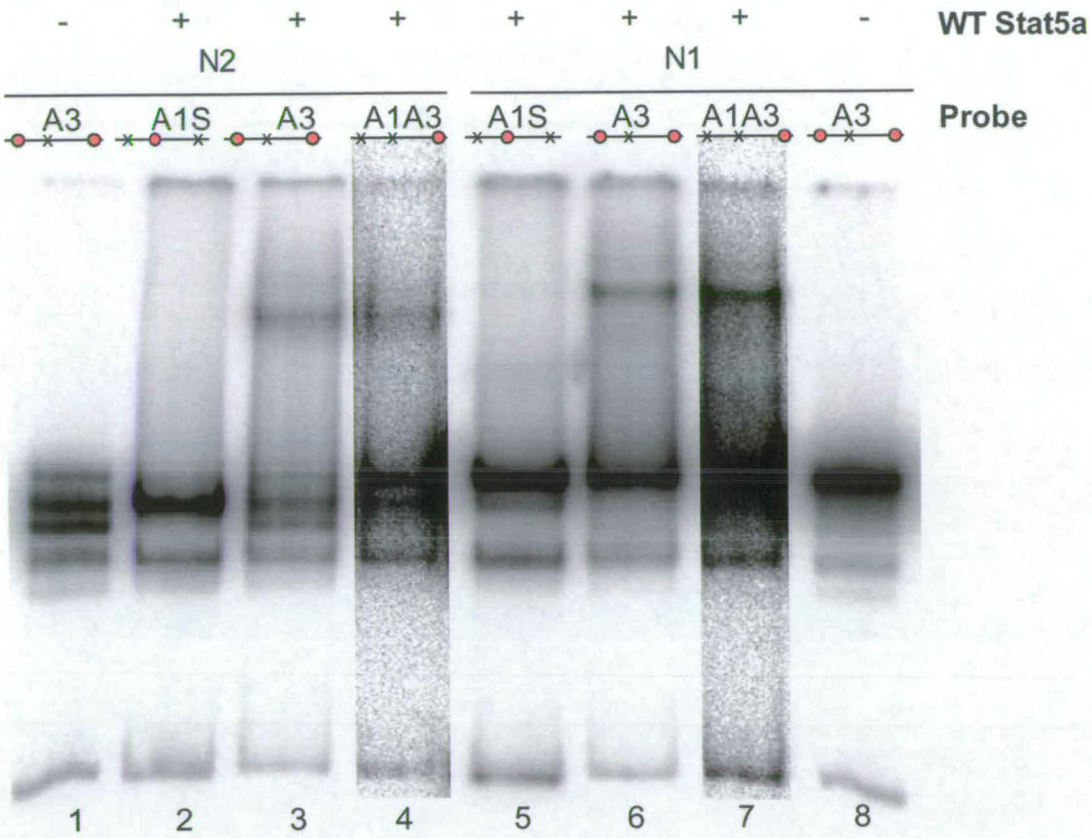


Figure 5.9 Stat5a Binding to Sites in Nucleosomal DNA.

WT Stat5a bandshifts were carried out on nucleosome isomers N1 and N2 on variations of nAB containing Stat5 binding site mutations indicated above lanes, along with a representation of the binding sites. Red circles represent WT binding sites, and black crosses represent mutated binding sites. The two A1A3 bandshifts are of a probe labelled with a lower specific activity than the other shifts, the apparent intensity of these lanes has been increased during Quantity One analysis although all lanes are from the same gel. Plus or minus signs about the gel indicate the presence or absence of WT Stat5a in each lane.

binding site is a maximum of 26bp from the boundary of the N3 nucleosome (calculated from the isomer at –234bp from the transcription start site) and a minimum of 15bp (from the isomer at –213bp). The A1 binding site is a weaker site than the A3 binding site (Burdon et al., 1994b) indicating that the difference in Stat5a binding observed is not due to binding site strength. Thus Stat5 cannot bind to a site 40bp from a nucleosome boundary, but it can bind to a site 15bp from the boundary.

The N3 nucleosome position is composed of a collection of positioned nucleosomes at –234, –225, –219 and –213. An end positioned nucleosome can be at either end of the fragment, thus it is possible that a fraction of the A1 binding sites are available through this way for binding in the N3 position, for example, if an end-positioned nucleosome was positioned at the downstream end of nAB. Alternatively, contaminating bands from the N1 and N2 positions in the N3 isolation may be responsible for this bandshift. There is no lane containing non Stat5a shifted position N3 on A3S to observe which band is being shifted, however a comparison of the proportion of signal in each nucleosome positioning band in N3 isolations from the other experiments shows that the percentage in position N3 on other versions of nAB is approximately 66.7% for a Stat5a bound lane. In lane 5 of Figure 5.10 the percentage of the signal in the N3 position is 73.5%. This is a 6.8 percentage-point difference. These values are different enough to suggest that Stat5a may not be binding to the N3 position, but rather to other contaminating bands in the isolation resulting in the higher proportion of signal in the N3 position, however this may also reflect a slightly different nucleosome distribution on A3S. In Figure 5.2 the N3 position on A3S is more prominent than in any other reconstitute on nAB. Binding to N3

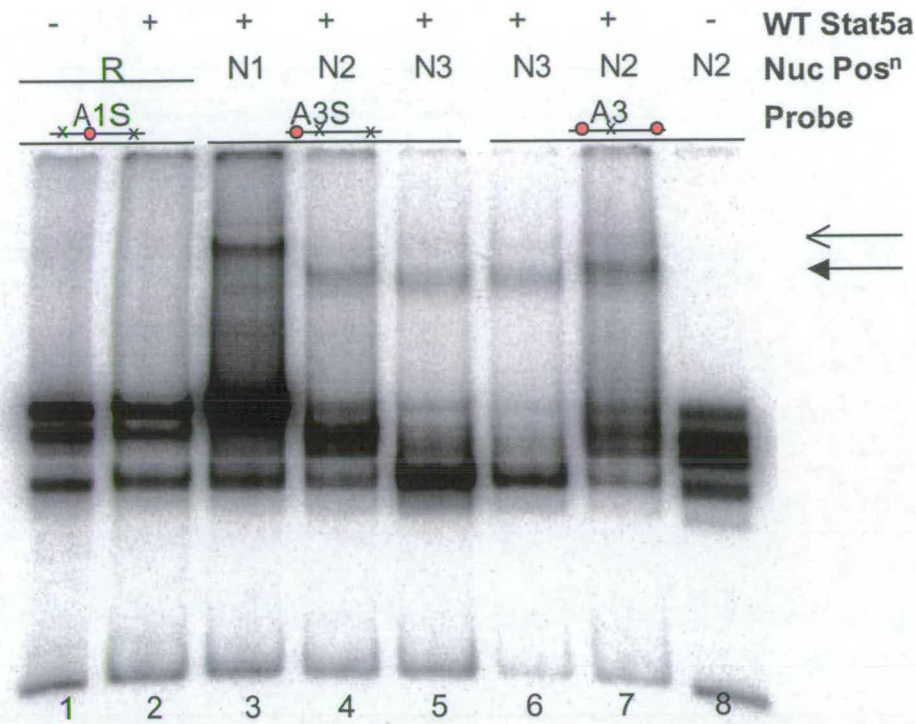


Figure 5.10 Stat5a Binding to Sites at the Nucleosome Boundary.

WT Stat5a bandshifts were carried out on isolated nucleosome positions N1, N2 and N3 and on unfractionated reconstitutes (R) on probes containing various Stat5 binding site mutations. The names of the mutated fragments, as described in Figure 4.13 are shown above the gel with a representation of the binding sites. Red circles represent WT Stat5 binding sites, and black crosses represent mutated binding sites. Plus or minus signs above the gel indicate the presence or absence of WT Stat5a in each lane. The isolated nucleosome position isomer that was used for each shift is also shown above each lane. Arrows to the right of the gel mark the secondary (open arrow) and primary (closed arrow) shifted bands.

cannot be ruled out as the A3S N3 band migrates at the same level as the A3 N3 band rather than the N2 or N1 bands (Figure 5.10 compare lane 5 to lane 6 and to lane 4 or 7) suggesting that binding is to N3. The possibility still exists that Stat5a can bind to sites in DNA at the boundary of a nucleosome.

Taken together the results in section 5.5 indicate that the position of a Stat5a binding site relative to a reconstituted nucleosome affects the binding of Stat5 to that site. Therefore it follows that the positions of nucleosomes over gene promoters must have a major role in the regulation of Stat5a responsive genes such as BLG.

5.6 Restriction Enzyme Access to Nucleosomal Binding Sites.

The mapped nucleosome positions indicate that the A1 binding site is covered by the nucleosome in the N3 nucleosome position. Stat5a access to A1 in the N3 position was questioned in 5.5.2 and the issue of access to this site remains unresolved. To investigate access to the A1 binding site in the N3 nucleosome position, restriction digests were carried out on DNA in all three nucleosome positions for comparison. The A1 mutation was designed to create an *Rsa*I cleavage site (Figure 4.1), an enzyme which does not cut elsewhere in the nAB sequence. Digests of a probe containing the A1 mutation with *Rsa*I will show restriction enzyme access to DNA specifically at the A1 site. *Rsa*I digests of reconstitutions onto the A1S probe show that there is partial access of *Rsa*I to the A1 site on the N3 position (Figure 5.11 lane 5) this is consistent with Stat5a being able to bind at this site. No DNA sequence in a nucleosome is completely protected, and access depends on the proximity of the sequence to the nucleosome boundary (Anderson et al., 2002). Sequences within a nucleosome can be protected by up to 1000 times compared to DNA, whereas for sequences near the nucleosome boundary

this falls to 50 times (Linxweiler and Horz, 1984) this is illustrated in (Figure 5.3). This can occur while the nucleosome remains positioned, probably by spontaneous partial unwinding of the DNA. It is likely that as the A1 Stat5a binding site is positioned near to the periphery of the N3 nucleosome, it allows Stat5a binding. The A3 Stat5a binding site, which is positioned more central to the nucleosome core does not allow Stat5a binding.

5.6.1 Non-Nucleosomal DNA Interactions with the N2 Positioned Nucleosome.

Interestingly, the *RsaI* restriction enzyme accessibility assay also reports that there is no access to the A1 binding site in the N2 nucleosome position. In fact the A1 site on the N2 nucleosome binds Stat5a well (Figure 5.10, lane 4). A repeat of the digest gave the same result. The N2 position, as mapped in Chapter 3 should place the A1 binding site external to the nucleosome core (Figures 3.17 and 5.1). Its position at –169 and –176 places the nucleosome boundary at a minimum of 22bp from the closest edge of the A1 Stat5a binding site. At this distance one would predict that the presence of a nucleosome should not inhibit restriction enzyme digestion although the behaviour of individual restriction endonucleases within a chromatin environment has not been studied. Exonuclease III digestion of the N2 position from the upstream end during nucleosome mapping revealed a collection of nucleosome specific pause sites up to 30bp external to the recognised nucleosome boundary (Figure 3.11). These pauses do not initiate a characteristic 10bp pattern and are only present at low concentrations of ExoIII and thus do not represent a true nucleosome boundary. The extra pauses were explained by interactions of non-nucleosomal DNA (i.e. DNA external to the 146bp in the histone core) lying on the surface of the

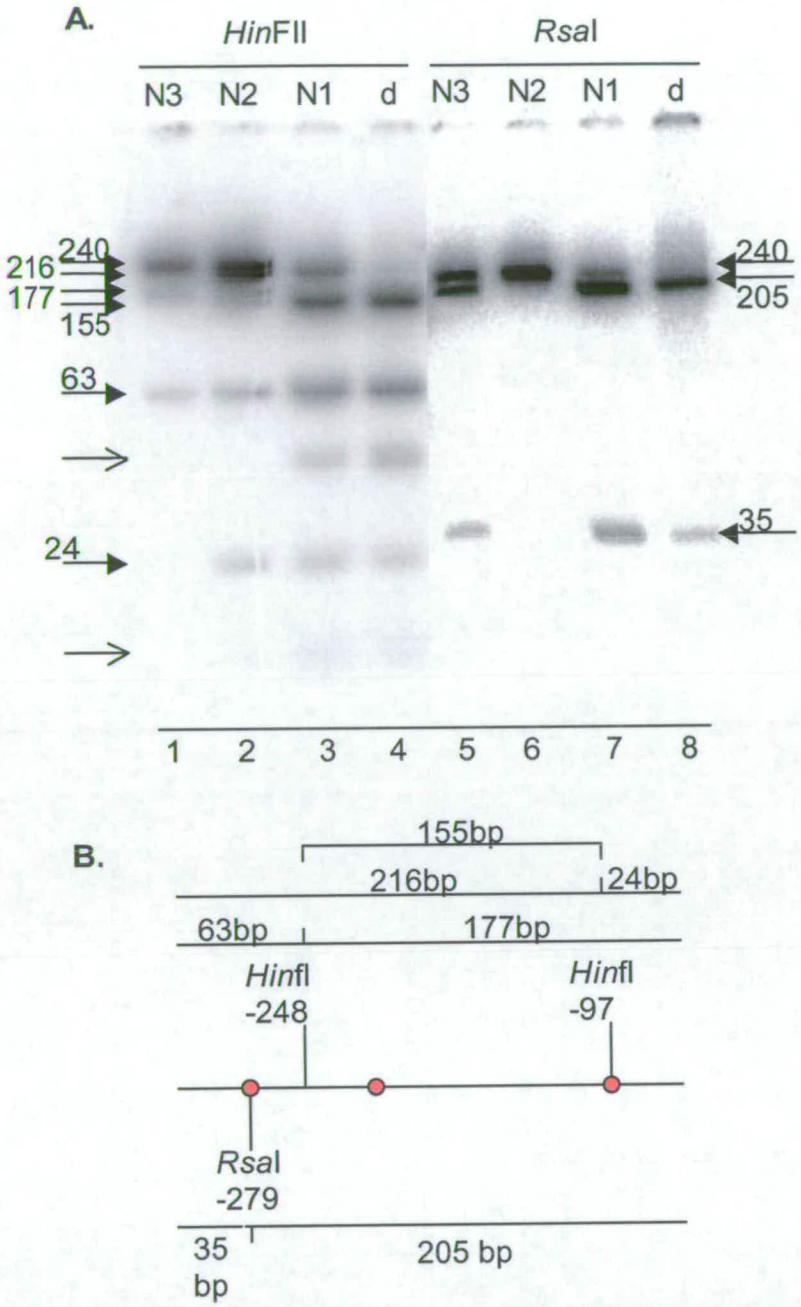


Figure 5.11 Restriction Enzyme Accessibility of Nucleosome Positioning Isomers.

Isolated nucleosome positioning isomers (N1, N2 and N3) and naked DNA (d) were digested by either *Rsa*I (A1S) or *Hin*fI (WT), and the resulting products separated on an 8% polyacrylamide gel in 1 X TBE, run at 200V constant for 3 hours (**A**). Digests were carried out in 0.1 X supplied buffers, to maintain low salt concentrations. Each contained $MgCl_2$ at 1mM. Solid arrows mark the predicted digest products, and open arrows mark the unexplained bands. Predicted fragment lengths are indicated on a representation of nAB showing the Stat5 binding sites as red circles (**B**).

nucleosome. It is entirely possible that this interaction also accounts for the inability of *Rsa*I to digest at the A1 Stat5a binding site in the N2 position Isomer, although both the ExoIII and *Rsa*I experiments were carried on nAB that contained the wild type A3 binding site, and do not contain an identifiably NX position.

5.6.2 *Hin*fI Digests.

Because of these unexpected results, restriction digests were also carried out with a second enzyme, *Hin*fI. Digestion sites and expected fragment sizes for this enzyme are illustrated in Figure 5.11B and relation of the cleavage sites to nucleosome positions in Figure 5.12. *Hin*fI cuts twice in nAB, at –248 and at –97bp. The –248 site is 30bp internal to the *Rsa*I A1 site and the –97bp site is in the StM Stat5a binding site. *Hin*fI digestion did not result in digestion to completion of any of the isolated nucleosome positions. This was expected as the two sites are 151 bp apart leaving room for one nucleosome with a dyad at –172 +/- 3bp. There is no corresponding nucleosome position on the nAB sequence although several are close. Each nucleosome position should therefore have at least one site where digestion is hindered by the presence of a nucleosome.

Of the three positions, DNA from the N1 position is digested most completely, with no uncut DNA detected. Products of 24 and 216 bp show that in N1 the downstream (–97bp) *Hin*fI site is available for digestion, and a band of 63bp indicates that the upstream (–248) *Hin*fI site also is available. The absence of a band of 177bp and a strong band at 155bp indicates that there is little single digestion at only the –248bp site, cleavage is either at the –97 alone or at both the –97 and the –248 sites. The strong N1 nucleosome isomer at –184bp overlaps the –248 *Hin*fI site by 9bp, and the –198

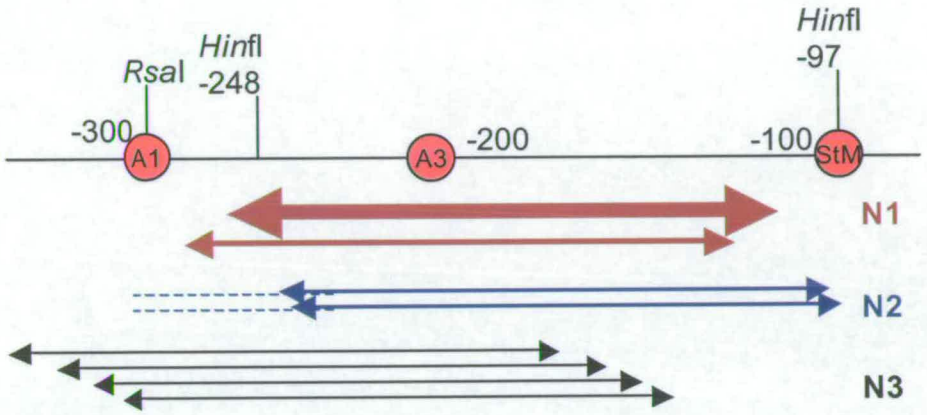


Figure 5.12 Restriction Sites in Relation to Mapped Nucleosome Positions.

This schematic diagram shows the cleavage sites of *Rsa*I and *Hin*fI in relation to the sequences protected by nucleosomes. The extra sequence interactions detected in N2 are marked by dashed lines.

nucleosome overlaps the same by 23bp (Figure 5.12). This cleavage pattern may represent a double digest of the more centrally placed –184 nucleosome and a single digest of the –198 nucleosome which places the upstream *Hin*fI cleavage site further within the nucleosome. A seventh unexplained band migrating about 50bp is present in both the N1 lane and the DNA lane, there is no corresponding *Hin*fI or similar restriction site present in the nAB sequence and the enzyme is not reported to have any star activity. The source of this band must be something in the sequence (perhaps introduced by a PCR mutation) that is covered by the nucleosome in positions N2 and N3 but not N1. A faint band migrating at approximately 10bp suggests that the 50bp band may be a cleavage product of the 63bp band.

The N2 position contains a significant proportion of DNA that is not digested at either *Hin*fI site demonstrated by a band of 240bp. But bands of 216, 177, 153, 63 and 24bp are all seen, indicating that all sites are partially accessible and that in a proportion of the nucleosomes both sites are cut (Figure 5.11). This reflects the multiple nucleosomes in each isolated position. The nucleosome at –169bp sits almost exactly between the two sites, just overlapping the –97 *Hin*fI site (Figure 5.12). The –176 nucleosome also sits between the two *Hin*fI sites (Figure 5.12), overlapping the –248 *Hin*fI recognition site by one bp. These positions would suggest that *Hin*fI should digest N2 more efficiently than N1. The only explanation for the under-cutting of sites with the N2 nucleosome position compared to the N1 is the previously described interactions of non-nucleosomal DNA with the N2 positioned nucleosome, blocking access to sites in this stretch of DNA.

The issue of the total failure of *Rsa*I to cut the N2 position, while the more central *Hin*fI site at –248 is partially cut (as demonstrated by the appearance of the 63bp band) remains unresolved. If the unexplained band is a digestion product of the 63bp *Hin*fI band, then *Hin*fI must also cut at the upstream end of this fragment, but not in N2. A possible explanations for the peculiar digestion of N2 is that it is caused by an end-positioned nucleosome. The N2 position cannot be an end position as the –279 *Rsa*I site is 35bp from the end of the fragment. It is unlikely that a nucleosome would be associated with this stretch of DNA, but not with the –248 *Hin*fI site 63bp from the end as this would mean less than one turn of DNA was wrapped round the nucleosome. An end positioned nucleosome is also predicted to migrate at the fastest rate (Pennings, 1997). The N2 position migrates more slowly than N3, suggesting that N3 is the end positioned nucleosome. Another, more likely explanation is that the nucleosome/DNA interactions external to the core DNA are involved, blocking access to sites in this sequence. Finally, there could be a higher order structure in the DNA sequence that does not allow *Rsa*I to cut, but this would be expected to be present in digests other than of the N2 nucleosome position.

The N3 nucleosome position digest contains neither a 24bp band nor a corresponding 216bp band, therefore is not cut at –97 by *Hin*fI. A fraction of the N3 DNA is cut at –248, but a significant proportion of probe is uncut. This suggests a nucleosome that is able to cover the –97 *Hin*fI site, but still protecting the –248 *Hin*fI site. This is hard to explain by the binding of a single nucleosome. The –97 *Hin*fI site is positioned over the StM Stat5 binding site, suggesting that StM may be covered by the nucleosome in the N3 position. The probe A1A3, which contains only this site, is still shifted in

the N3 nucleosome position (Figures 5.7 and 5.8). If as the restriction data suggest an end positioned nucleosome sits at the downstream end of the probe this shows that for StM at least Stat5 can bind to DNA near the edge of a positioned nucleosome.

The restriction enzyme accessibility data backs up the hypothesis that there is partial although unpredictable access to DNA at sites close to the boundary of a positioned nucleosome on nAB. The A3 binding site that does not allow Stat5a binding is positioned more toward the nucleosome dyad than the A1 binding site is, A1 may still allow Stat5a binding although it is in nucleosome associated DNA. The position of the Stat5a binding site on the nucleosome may be crucial to whether or not Stat5a will bind to a weaker binding site.

5.7 Do Stat5a Dimers make Higher Order Interactions on nAB.

My hypothesis is that the wrapping of DNA round a positioned nucleosome in the N1 position brings together the strong StM and the weaker A1 Stat5a binding sites, potentially facilitating DNA bound Stat5a-Stat5a dimer-dimer interactions, otherwise referred to as tetramerisation. If Stat5a dimers interact with each other as a tetramer on a nucleosome associated nAB probe their synergistic interaction should exhibit a more stable Stat5a binding to the A1 site, and hence more stable total Stat5a binding to the probe (John et al., 1999; Vinkemeier et al., 1996; Vinkemeier et al., 1998). Usually consensus and nonconsensus binding sites must be placed close to each other in cis on the DNA sequence to facilitate tetramerisation. Organisation of DNA into a nucleosome could have the same result, bringing two Stat5 binding sites close enough to allow tetramerisation.

5.7.1 N1 Competitor Studies with A3.

The relative strength of both WT Stat5a and Stat5aW37A binding to the N1 nucleosome position on a WT probe was assessed using competition studies with an unlabelled double stranded oligonucleotide to the A3 binding site as competitor. The A3 site was chosen as it binds Stat5a weakly, and thus should be able to reveal more subtle changes in the binding affinities of the two Stat5a species to the nucleosomal probe. Bandshift reactions were carried out as previously described, with the exception that the final incubation at room temperature was for 10 minutes before addition of the competitor and incubation for a further 10 minutes at room temperature before the reaction was loaded onto a polyacrylamide gel in TBE.

The two competitor concentrations of 20- and 200-fold molar excess (relative to the labelled probe) were chosen to give a maximal range of competition. A3 almost completely competes Stat5 binding to the StM oligonucleotide probe at 200-fold excess, and as expected a significant amount of signal remains at 20-fold excess (Figure 4.2.B).

The primary and secondary shifted bands are seen for both WT Stat5a and for Stat5aW37A shifts in non-competed lanes, Figure 5.13. The secondary band is totally competed in W37A shifts with a 20-fold excess of the A3 oligonucleotide, but in WT Stat5a shifts it is still visible at 20-fold. The percentage remaining signal after competition in the combined primary and secondary shifted bands was calculated as a percentage of the signal in the non-competed lane, after normalisation to the total signal in each lane (Quantity One analysis of a phosphorimage). These values are listed below each lane in Figure 5.13A. The data suggests that WT Stat5a may have a subtly higher affinity for the N1 probe than Stat5aW37A does, but the values

Figure 5.13 A3 Competition on the N1 Position at 20 and 200-fold Molar Excess.

The strength of binding of WT Stat5a and Stat5aW37A to the N1 nucleosome position was studied using competition by the weak Stat5 binding site A3. An image of the gel is shown in **A**, and a graphical representation of the results after analysis of a phosphorimage in **B**. Arrows mark the secondary and primary shifted bands. The percentage remaining signal in the shifted band and the lane number are marked on the gel. Above the gel the Stat5a used to shift in each lane and the fold molar excess of competitor (relative to the labelled probe) are indicated.

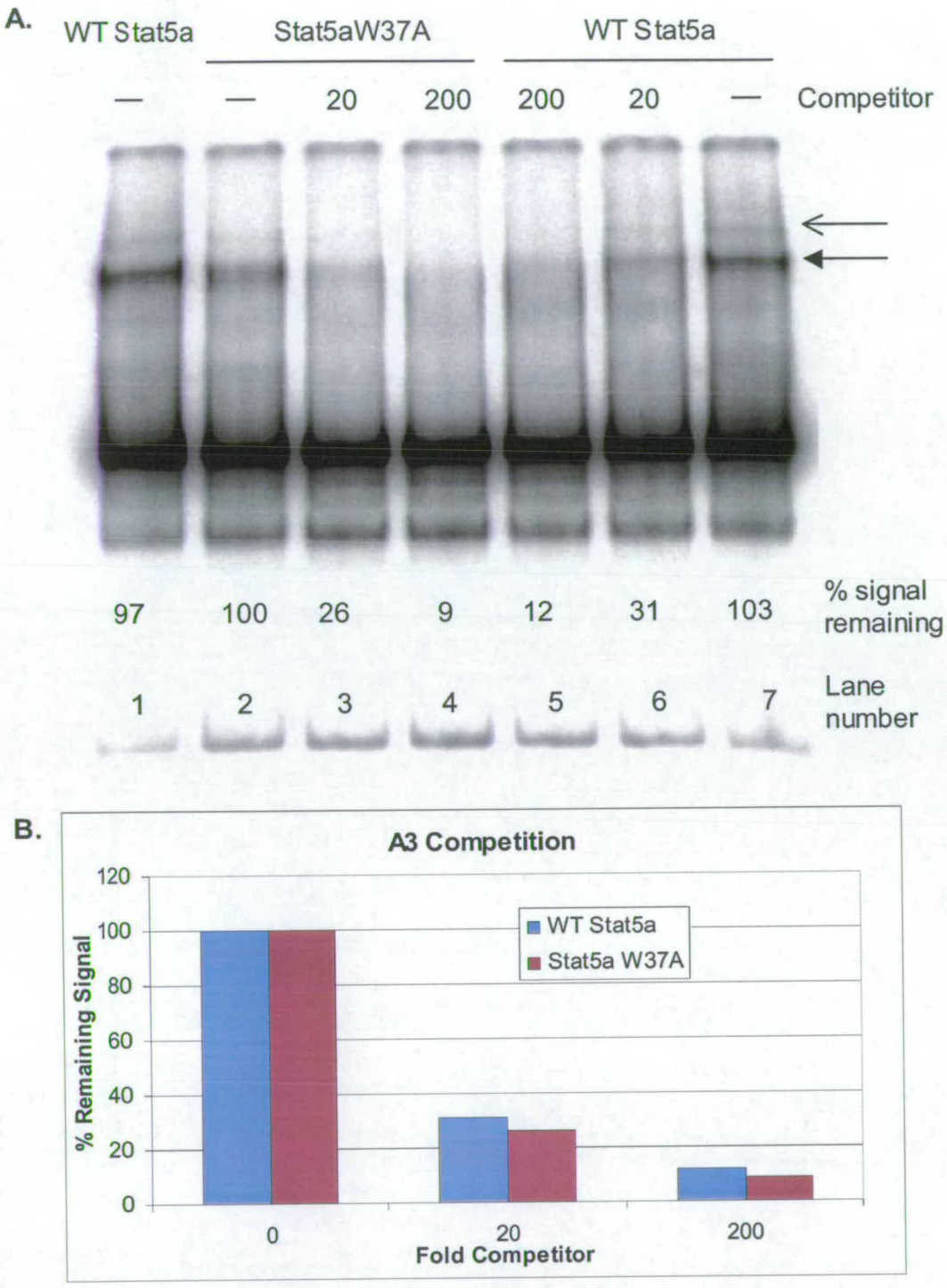

A.
WT Stat5a
Stat5aW37A
WT Stat5a
— — 20 200 200 20 — Competitor
97 100 26 9 12 31 103 % signal remaining
1 2 3 4 5 6 7 Lane number
B.
A3 Competition
% Remaining Signal
120
100
80
60
40
20
0
WT Stat5a
Stat5a W37A
0
20
200
Fold Competitor

Figure 5.13.

are probably not different enough to indicate a real difference. These results are also shown as a graph in Figure 5.13B.

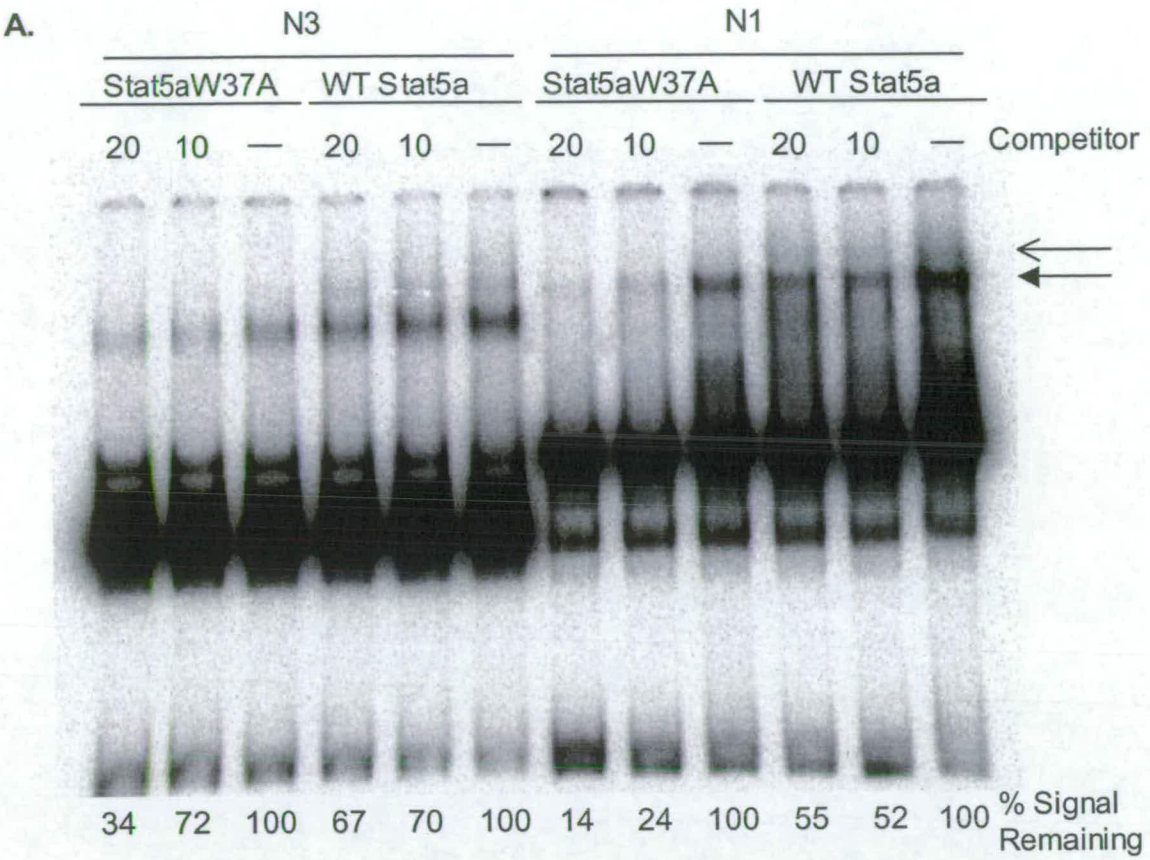
5.7.2 N1 and N3 Competitor Studies with A3.

The competition experiment was repeated on nucleosome positions N1 and N3 on a WT probe. N3, unlike N1, is predicted to cover a second Stat5a binding site that is potentially involved in tetramerisation. From the nucleosome mapping this is predicted to be A1, although the restriction enzyme analysis experiments suggest that there is access to this site, and that it may be the StM site that is covered by the N3 nucleosome. Stat5a does bind to StM in the N3 nucleosome position (Figures 5.5 and 5.8), and probably also to A1 as demonstrated in Figure 5.10. The N2 position was not used because of its unexplained behaviour; any results could not be interpreted in the context of nucleosome position. A lower competitor range was chosen for this experiment to examine any more subtle changes in Stat5a affinity for nAB. Results of competitor experiments with 10 and 20-fold molar excess of unlabelled A3 are shown in Figure 5.14.

For both the nucleosome positions N1 and N3, there is no difference between the addition of 10 or 20-fold molar excess competitor on shifts with WT Stat5a (Figure 5.14). In position N1, 70% of the shifted signal remains after competition at a 10-fold and 67% remains after competition at a 20-fold excess for ten minutes. 52% and 55% signal remain in Position N3. On the other hand competitions of Stat5aW37A shifts on position N1 demonstrate a 1.7 fold difference between 10 and 20-fold excess competitor (72% remaining signal for 10-fold and 34% for 20-fold excess). The same difference for N3 is 2.1 fold (24% and 14% remaining). These results again suggest that WT Stat5a may have a higher affinity in general for a nucleosomal probe than

Figure 5.14 A3 Competition of Stat5a Binding to N1 and N3 at 10 and 20-Fold Molar Excess.

The strength of binding of WT Stat5a and Stat5aW37A to the N1 and N3 nucleosome positions was studied using competition by the weak Stat5 binding site A3 at a lower fold excess competitor to probe more subtle differences. An image of the gel is shown in **A**, and a graphical representation of the results after analysis of a phosphorimage in **B**. Arrows mark the secondary and primary shifted bands. The signal remaining in the shifted bands expressed as a percentage of the uncompeted shifted signal for the appropriate Stat5a nucleosome position is marked below the gel. Above the gel the Stat5a used to shift in each lane, the fold molar excess of competitor (relative to the labelled probe in each reaction) and the shifted nucleosome position are marked.


A.
N3
N1
Stat5aW37A
WT Stat5a
Stat5aW37A
WT Stat5a
20 10 — 20 10 — 20 10 — 20 10 — Competitor
34 72 100 67 70 100 14 24 100 55 52 100 % Signal Remaining

B.

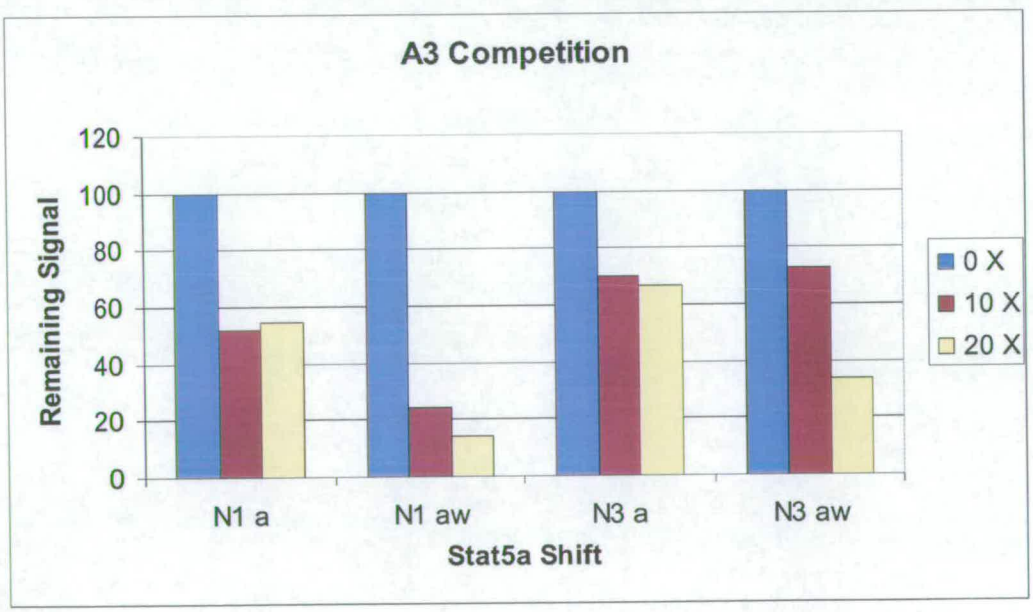

A3 Competition
Remaining Signal
120
100
80
60
40
20
0
N1 a
N1 aw
N3 a
N3 aw
Stat5a Shift
0 X
10 X
20 X

Figure 5.14.

Stat5aW37A does. WT Stat5a shows no difference between a 10 or a 20-fold competitor, perhaps reflecting a plateau in the stability of the Stat5a complex, which is overcome at a higher fold competitor such as the 200-fold excess in Figure 5.13. It is possible this is mediated by the tetramerisation interaction as this is not observed with the W37A mutant. These results are shown in graphical form in Figure 5.14.B. As Stat5a can bind to both the A1 and StM sites in both nucleosome positions, these subtle differences in the position of the nucleosome may not be important. However a nucleosome that places the Stat5 binding site more centrally, as for the A3, would be predicted to be more likely to have an effect.

When the relative affinities of Stat5a for the N3 and the N1 nucleosome positions were compared, a surprising result emerged. N3 positions in general appeared to have a greater resistance to competition by A3 than N1 positions did. For WT Stat5a, N3/N1 ratios for 10 and 20-fold molar excess are 1.35 and 1.22 respectively, and for Stat5a W37A they are 3.00 and 2.43 respectively. It may be that the arrangement of DNA on the N3 nucleosome is more favourable to Stat5 binding than the N1, as the StM binding site is further from the nucleosome. The N3 shifted bands are very weak making any quantification unreliable, so no solid conclusions can be drawn from this. The angle of the DNA exiting the nucleosome or the rotational setting, i.e. whether or not the DNA sequence is orientated correctly to allow interactions would both influence Stat5a binding interactions. The rotational setting will particularly influence binding to A1 and A3 (see section 6.2.1). DNaseI experiments to determine the rotational setting on these nucleosomes are in progress. The difference between the affinities of WT Stat5a and Stat5aW37A for N3 is similar to those for N1 but the N3 position has a

slightly higher difference of 2.1 compared to 1.7 fold for the N1. This theory requires further testing, as it backs up the results in Figure 5.8, which suggested that nAB might shift at the secondary shift level more in the N3 position than in N1 or N2.

5.8 Chapter Conclusions.

The work in this chapter has shown that Stat5a will bind to a reconstituted chromatin probe, and to each individual nucleosome position isolated from this. Stat5a binds to the position isomers in order N2/NX>N1>N3. Binding to individual Stat5 binding sites inside nucleosome associated DNA is also described. Stat5a will not bind to the A3 binding site, which is situated in the middle of a nucleosome. The question of whether or not Stat5a can bind to the A1 Stat5 binding site at the periphery of a bound nucleosome has not been fully answered, but restriction analysis suggests that partial access is available to this site when it is near the edge of a nucleosome. There is no restriction enzyme access to the StM site in position N3, indicating that it may be covered by an end positioned nucleosome, yet Stat5a will bind to this site. The position of a binding site within nucleosomal DNA appears to have a major role in whether or not Stat5a will bind. Alternatively, as the StM binding site is much stronger that the A3 and the A1, binding to this site may be a reflection of the affinity Stat5a has for the sequence. This could be investigated further by substituting the A3 site for an StM site, assuming that this does not cause further changes in nucleosome positioning.

This chapter also described the anomalous migration of the N2 nucleosome position as a secondary band, termed NX. It is likely that this is a result of a change in the stability of the N2 band that causes an already present structural abnormality in this nucleosome (the restriction of access to the

upstream sequences) to alter the migration of N2//NX through a non-denaturing polyacrylamide gel; perhaps by altering the likelihood that a nucleosome positions to the N2 or the NX site. Alternatively N2 and NX could position at the same location and the extra contacts in NX may cause a different migration.

Finally Stat5a dimer-dimer, or tetramerisation interactions were studied. A difference is detected between the affinity of WT Stat5a (which can make tetramerisation interactions) and Stat5aW37A (which cannot) for nucleosomal probes especially at low molar excess of a weak competitor. Further study must be carried out to characterise this further.

6 DISCUSSION.

It is clear that the chromatin structure over the BLG gene plays a major role in its regulation, most likely by regulating the accessibility for transcription factors such as Stat5. In this thesis I have demonstrated that recombinant Stat5a will bind to reconstituted nucleosomes *in vitro*, and that it will do so on Stat5 binding sites that are located within the outer 25bp of core protected DNA. Stat5a will not bind to sites internally located by 40bp or more from the nucleosome boundary. This information forms the basis for resolving the precise role of a positioned nucleosome on the BLG promoter.

6.1 Stat5 Binding to DNA.

The binding of both recombinant Stat5, produced in a baculovirus expression system, and Stat5 from a nuclear extract to binding sites in the 240bp fragment of the BLG promoter (nAB) and to oligonucleotide probes corresponding to the individual Stat5 binding sites has been studied in Chapter 4. Both sets of shifts exhibited multiple shifted bands (Figures 4.2 and 4.10). Multiple bands were present also in shifts using Stat5 containing the W37A mutation, which has been reported previously to block the formation of Stat5 dimer-dimer interactions, known as tetramerisation (John et al., 1999; Vinkemeier et al., 1998), indicating that the multiple shifted bands observed in probes shifted by recombinant Stat5 are not due to tetramerisation interactions (but see section 6.1.1). The pattern of bands observed in shifts by recombinant Stat5 and Stat5 from a mammary gland nuclear extract are different, the most rapidly migrating shifted band migrates at the same level in both but there are three bands in rStat5 shifts, and only two in nuclear extract shifts.

An explanation for the appearance of multiple shifted bands is that a percentage of the Stat5 population is post translationally modified. Stat proteins are reported to be phosphorylated on residues other than the conserved tyrosine essential for activation of the protein (Beuvink et al., 2000; Park et al., 2001; Ridderstrale and Groop, 2001; Yamashita et al., 2001). Both Stat5a and Stat5b have been reported to be phosphorylated on serine residues (S725 and S779 in mouse Stat5a) (Yamashita et al., 2001).

The recombinant Stat5 produced in the baculovirus expression system was found to be phosphorylated on the conserved tyrosine (Y694 in Stat5a) and was able to bind DNA without the need to co infect cells with a JAK encoding baculovirus. This indiscriminate phosphorylation may also result in the inappropriate modification of other residues, which could affect the migration of the shifted complex through a polyacrylamide gel. Altering the conditions used to separate the shifted complexes resulted in the migration of shifted complexes as a discrete band (Figure 4.12).

6.1.1 Tetramerisation.

Recent work has suggested that the previously characterised N-terminal interactions involved in tetramerisation are not correct. The actual tetramerisation interface does not involve W37, but rather hydrophobic interactions between E29 and L78 and also between L15 and L77 (this numbering relates to Stat4) (Chen et al., 2002). These residues are all at least semi-conserved between Stats 1, 4, 5a and 5b (Figure 6.1). However the W37A mutant has been shown to prevent tetramerisation occurring, and even if it is not the amino acid that makes contact between two Stat5 dimers

A.

```
Stat5a          MAGWIQAQQLQGDALRQMQVLYGQHFPIEVRHYLAQWIESQPWDAIDLDNPQDRAQATQL
Stat5b          MAVWIQAQQLQGEALHQMQALYGQHFPIEVRHYLSQWIESQAWDSVDLDNPQENIKATQL
Stat4           MSQWNQVQQLEIKFLEQVDQFYDDNFPMEIRHLLAQWIENQDWEAASNNE----TMATIL
                *: * *.***: . *.*:: :*.::**:*:** *:****.* *:: . ::      ** *

Stat5a          LEGLVQELQKKAEHQVGEDGFLLKIKLGHYATQLQKTYDRCPLELVRCIRHILYNEQRLV
Stat5b          LEGLVQELQKKAEHQVGEDGFLLKIKLGHYATQLQNTYDRCPMELVRCIRHILYNEQRLV
Stat4           LQNLLIQLDEQLGRVSKEKNLLLIHNLKRIRKVLQGKFHGNPMHVAVVISNCLREERRIL
                *:.*: :*:::  :   *..:**  :* :  . ** .:.  *:.:.  * : * :*:*::
```

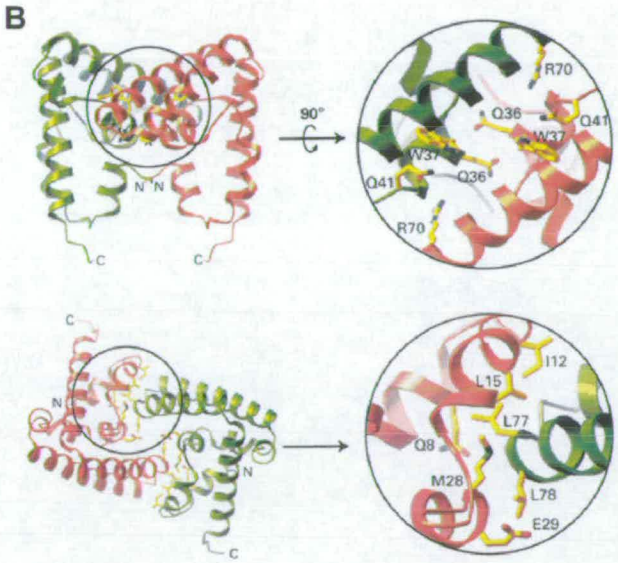


Figure 6.1 Interactions at the N-terminal Domains of Stats.

A. Alignments of Stats 4, 5a and 5b were carried out using ClustalW. Residues thought to be involved in tetramerisation are highlighted. Residues not fully conserved in Stat5 are in coloured font instead of highlighted. **B**. Top two images are the original model for Stat tetramerisation (Vinkemeier et al., 1998), the bottom two images are the revised model (Chen et al., 2002). The views are from different angles. Residues are highlighted in **A** blue for the Vinkemeier model and Yellow for the Chen model. The Structures in part B were taken from (Chen et al., 2002) and the numbering relates to Stat4.

the mutation still functions, probably by destabilising the N-terminal domain. The residues proposed to make the tetramerisation interactions are conserved between Stat5a and Stat5b, so these cannot explain why Stat5b does not make tetramerisation interactions as efficiently as Stat5a does. An

alignment of Stat5a and Stat5b reveals that most of the differences occur at the N- and C-terminal domains (see appendices). Residues within these areas are likely to be involved in tetramerisation and other protein-protein interactions, and may hold the key to the different activation profiles of the two proteins.

No role for tetramerisation has been either proven or disproved by the work discussed in this thesis, but several avenues of investigation remain open. Most significantly competition analysis has not been carried out on naked DNA to compare the binding strengths of the two Stat5a species to this, or on the N2 nucleosome position, which was shown to bind Stat5a better than the other nucleosome positions. It may also be useful to repeat the whole experiment using the newly identified set of N-terminal tetramerisation interactions (Chen et al., 2002).

6.2 Nucleosome Positioning on the BLG Promoter.

Nucleosomes position at precise locations on nAB, a fragment of the proximal BLG promoter. These migrate as three discrete bands when a reconstitute is separated by non-denaturing polyacrylamide gel electrophoresis (N1, N2 and N3 in Figure 3.5). Nucleosome position isomers that migrate as the N1 band are located most closely to the centre of nAB, with nucleosome dyads positioned at –184 and –198 bp from the transcription start site. The N2 positioning isomer is composed of nucleosomes positioned at –169 and –176 and the N3 position is a combination of end positioned nucleosomes and sequence directed nucleosomes at –234, –225, –219 and –213 bp (Figure 3.17). These groups of positioned nucleosomes were isolated and the binding of Stat5a to each group studied. Results demonstrated that Stat5a will bind to all three

nucleosome isomers at sites external to, or near to, the nucleosome boundary, but will not bind to sites located more centrally on the nucleosome.

6.2.1 Rotational Positioning on nAB.

The nucleosome positioning isomers on nAB do not appear to be rotationally related, as their positions on nAB do not exhibit a 10bp periodicity (Figure 6.2). However digestion with ExoIII results in a characteristic 10bp periodicity of the nucleosome specific ExoIII pause sites detected in the N1 and N2 isolates (Figures 3.11 and 3.12). This arrangement suggests that rotationally related nucleosomes do exist within each group of nucleosome position isomers. The pauses in N1 also align with those in N2, indicating that these groups of positions are rotationally related to each other (Figure 3.11). It is unclear why when there is evidence for the presence of rotationally positioned nucleosomes, the mapped positions do not agree with this.

Further analysis of the data suggests an explanation for this, the three positions mapped by ExoIII are located with dyads at –219, –198 and –169 nt from the transcription start site. These positions are separated by 19 and 31bp and thus are rotationally related with an approximate 10bp periodicity (Figure 6.2). The DNA protected by the nucleosome from ExoIII digest was calculated to be 149 bp, this is slightly larger than the accepted 146bp, indicating a small potential error in these maps. The nucleosomes mapped by core particle restriction digests are located with dyads mapping at –234, –225, –213, –184, –176 and –169nt from the transcription start site (Figures 3.16 and 6.2). With the exception of the nucleosome position at –169, which was detected only weakly, these nucleosome positions are also related with

an approximate 10bp periodicity, but this is out by 5bp from the periodicity detected by ExoIII digestion (Figure 6.2).

This difference in the periodicity is likely to be a reflection of the different mapping techniques used to calculate these positions. Nucleosome positions mapped over the beta globin gene displayed a similar discrepancy between the ExoIII mapped position, which detected a nucleosome with boundaries at –207 and –60 nt from the transcription start site (Kefalas et al., 1988), and a core DNA restriction mapped position with nucleosome boundaries at –204

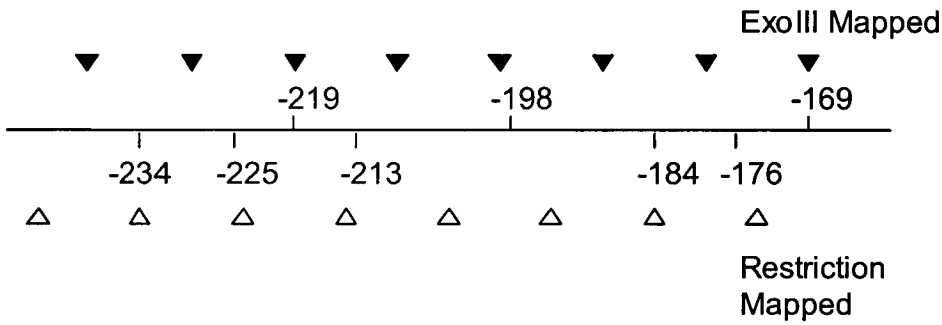


Figure 6.2 Rotational Setting of Mapped Nucleosome Positions.
The rotational setting of the nucleosome positions mapped on nAB is indicated in this schematic diagram. ExoIII mapped locations of the nucleosome dyad are indicated above the line, and restriction mapped dyad locations below the line. The ExoIII 10bp periodicity is marked by black triangles at –179, –169 etc and the restriction 10bp periodicity by white triangles at –234, –224 etc.

and –56 (Yenidunya et al., 1994). A similar discrepancy was observed with these two mapping techniques on the 5S rRNA gene (Buttinelli et al., 1993). This may be a general feature of ExoIII and core restriction mapping, and may reflect the degree of access each enzyme requires at the nucleosome boundary or the relation of cleavage sites to binding sites. A similar

comparison of the access of DNaseI and ExoIII to a positioned nucleosome showed that ExoIII consistently cut further into the nucleosome core than DNaseI (Prunell, 1983). These two enzymes also produced slightly different positions for nucleosome boundaries on the 5S rRNA gene (Meersseman et al., 1991; Simpson and Stafford, 1983; Pennings et al., 1991). MNase has also been demonstrated to produce core particles with slightly recessed 5′ ends (Cousins et al., 2004). The similar discrepancy of 4bp between the ExoIII and restriction digest mapping techniques in the beta globin map, along with the accepted errors in mapping, is sufficient to explain the alternative periodicity that results from the two mapping techniques on nAB, indicating that as suggested by the 10bp periodicity detected in the ExoIII digest, all the positions are rotationally related. A DNaseI digest of the unfractionated reconstitute would confirm if this is the case. Not enough is known about the mechanisms of ExoIII digestion within a nucleosome to be able to use this to determine the rotational setting of nAB on these nucleosomes, but a DNaseI digest may be more informative. Furthermore a DNaseI protection assay may also be able to detect Stat5 binding to each of the three binding sites on a positioned nucleosome.

The orientation of a Stat5 binding site against the nucleosome may be a factor in Stat5 recognition of a binding site that is located in a nucleosome, particularly as regards the A1 and the A3 Stat5 binding sites, both of which are only half consensus sites. The TTC of A1 and the GAA of A3 are separated by 76 bp, thus placing them at opposite sides of the DNA helix. The A3 site did not allow Stat5a binding, but the A1 binding site did. This may be due to the orientation of each binding site on the nucleosome or alternatively it may be due to reduced accessibility at sites further from a

nucleosome boundary. The function of rotationally positioned nucleosomes on nAB could be to ensure that binding sites for transcription factors such as Stat5, which may be located within the nucleosome, remain in a position that allows them to be recognised and bound. Different rotationally positioned nucleosomes may therefore favour binding to either the A1 or the A3 binding sites.

6.2.2 Mapping Nucleosome Dyads.

A degree of ambiguity remains as to exactly where on nAB nucleosomes are positioned. The strength of the interaction of DNA with the nucleosome boundary clearly varies between nucleosomes, for example N2 makes strong interactions at the upstream nucleosome boundary that are detected in ExoIII digests (Figure 3.11). Such differences affect the detection of nucleosome boundaries by nucleases, as the more tightly DNA is bound to the nucleosome the less likely the DNA will be made available to a nuclease (Figure 6.3).

A technique has been developed where nucleosome positions can be mapped by site-directed cleavage by hydroxyl radicals engineered into the nucleosome core near to the dyad axis (Flaus et al., 1996). Repeating the nucleosome map on nAB using this technique would allow nucleosome positions to be mapped absolutely with respect to the dyad axis rather than to nucleosome boundaries. Nucleosome positions on nAB have been mapped accurately using the ExoIII and core restriction techniques described in Chapter 3. These maps also provide additional useful information about the nucleosome boundary that would not be evident from a hydroxyl radical map. A combination of these techniques would allow nucleosome positions on nAB to be mapped beyond doubt.

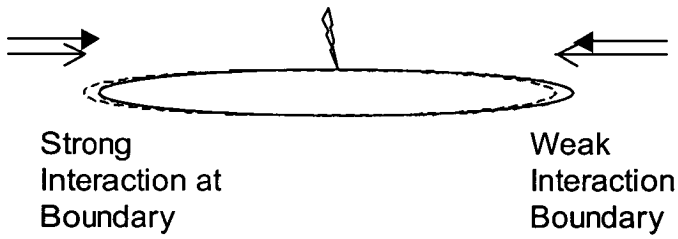


Figure 6.3 Mapping Nucleosome Boundaries.

The potential downside of nucleosome mapping techniques that use detection of nucleosome boundaries are highlighted in Figure 6.3. A weak interaction between the histone core and DNA, indicated at the downstream boundary of the nucleosome in this schematic diagram, may cause the nucleosome boundary to be detected further in to the core than it actually is (an open arrow marks the false boundary, and a closed arrow the true boundary). Conversely, an interaction of DNA external to the nucleosome with the histone core, depicted at the upstream boundary of the nucleosome in this diagram, may cause the nucleosome boundary to be detected further out from the core than it actually is. Again an open arrow marks the false boundary and a closed arrow the true boundary. This diagram depicts a single nucleosome, the dyad of which does not move, a technique that maps from the nucleosome dyad (lightning bolt) would not be affected by these boundary effects.

6.2.3 Implications for *In Vivo.*

The question of which of these nucleosome positions, if any, is present over the active gene promoter *in vivo* has yet to be addressed. The original *in vivo* map showed that an array of strongly positioned nucleosomes ends just upstream of the proximal promoter region studied in this thesis (Figure 3.1 and Boa, 1999).

Protected DNA bands are still detected in the *in vivo* nucleosome map of the proximal BLG promoter region (incorporating nAB) in mammary tissue (Figure 3.1), which suggests that a nucleosome remains at this location. This

may reflect the proportion of cells in the mammary gland that are not expressing BLG, or alternatively it may reflect the presence of a different, perhaps remodelled, nucleosome structure that does not protect the DNA from digestion to the same extent. One proposed mechanism for the action of nucleosome remodelling complexes, such as SWI/SNF, is the removal of H2A and H2B leaving only the $(H3H4)_2$ tetramer bound to DNA (Workman and Kingston, 1998; Bruno et al., 2003). NF1, which acts synergistically with both Stat5 and GR for maximal expression of WAP, has been shown to be able to bind to such a structure but not to a complete nucleosome octamer, this interaction is mediated by GR binding to the nucleosome (Mukhopadhyay et al., 2001; Spangenberg et al., 1998). There are binding sites for GR, NF-1 and Stat5 in the proximal BLG promoter (Figure 1.8 and Figure 6.4). A similar mechanism may take place on the BLG promoter, whereby the binding of GR and also Stat5 results in a change in the chromatin structure facilitating the initiation of transcription from the promoter.

It is a feature of active genes that nucleosomes appear to be displaced from the immediate promoter, usually after binding of transcription factors that can recognise their binding sites within the nucleosome (Cosma, 2002; Boeger et al., 2003; Lee et al., 2004). If this is the case on BLG, the role of a positioned nucleosome on the proximal promoter would be confined to initiating the formation of the transcription complex.

6.2.4 A Poised Chromatin Structure?

The *in vivo* map determined the chromatin structure over both the active BLG gene, in lactating mammary gland, and the chromatin structure in liver, a tissue where BLG is never activated. Two of the DNaseI hypersensitive sites

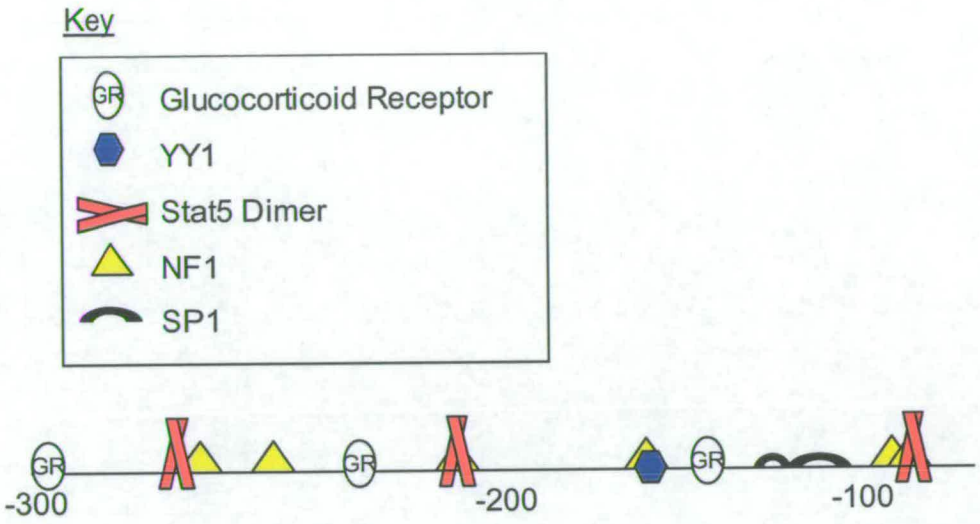


Figure 6.4 Transcription Factor Binding Sites on nAB.

Schematic diagram showing selected transcription factor binding sites on fragment nAB of the BLG promoter. The box above the sequence contains the key to which shape represents each factor. Figure 1.8 shows the sequences of each transcription factor binding site.

observed over the BLG gene, HSIV and HSV, are present up to the onset of milk protein production in mammary tissue and disappear with the upregulation of milk protein gene expression (Whitelaw and Webster, 1998). These are not present in liver suggesting that non-lactating mammary tissue may contain a different chromatin structure again from that observed in either the transcribed or the repressed genes (Figure 6.5).

The function of such a poised chromatin structure could be to maintain the BLG promoter in a mammary specific chromatin state that is ready to be activated during late pregnancy and lactation. Mammary gland specificity is conferred by the 406bp proximal promoter region of BLG in transgenic mice, although the specific mechanism of this is unknown (Webster et al., 1995). It

is possible that a specific chromatin structure over the proximal promoter, developed during mammary gland development, is responsible for the mammary specificity of BLG and results in a specific Stat5 binding pattern. This could explain why all three Stat5 binding sites are essential for maximum gene activity in mammary cells, but not in CHO cells (Demmer et al., 1995; Burdon et al., 1994b).

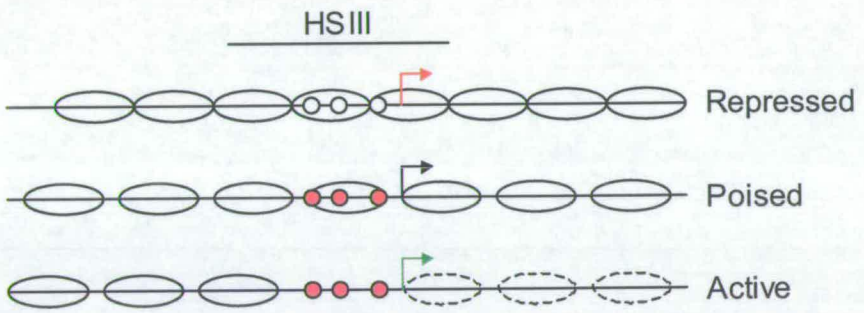


Figure 6.5 *In vivo* Chromatin Structures over BLG.

A schematic diagram of how the chromatin structure over the BLG promoter may look. The repressed structure is that seen in liver, a tissue that never expresses BLG. The poised structure may contain a positioned nucleosome over the proximal promoter such as those detected in the *in vitro* map, this nucleosome is removed in the active promoter, and the nucleosomes throughout the gene are disrupted.

A poised chromatin state may involve the positioning of a nucleosome at a specific location over the proximal promoter. For this reason it would be interesting to study the chromatin structure over the BLG promoter in mammary gland at other developmental stages, for example in virgin or involuting mammary gland. It would also be helpful to carry out this mapping with a higher resolution technique such as ligation mediated PCR (LM PCR) which would enable the nucleosome position *in vivo* to be mapped

to the base pair. Nucleosome positions observed *in vitro* do not necessarily occur *in vivo* as other factors, such as proteins bound to DNA, may influence nucleosome position.

Further studies of the *in vivo* structure during mammary gland development could determine which of the Stat5 activation models proposed below are relevant. Chromatin immunoprecipitation (ChIP) could be utilised to detect the timepoint when Stat5 is first detected bound to the BLG promoter and the chromatin structure compared with this. Another Stat family member, Stat3, is involved in mammary gland involution and can also bind to the Stat5 binding sites present in nAB. The appearance of Stat3 could also be monitored by the same methods. The transcription factor ying yang 1 (YY1), a mammalian polycomb group protein, and a second unidentified factor have been demonstrated to repress the β-casein promoter. Binding of Stat5 to the promoter overcomes YY1 mediated repression (Lee and Oka, 1992; Meier and Groner, 1994). A binding site for YY1 is present in the proximal BLG promoter region. YY1 may also have a role in regulation of the BLG gene.

6.2.5 A Role for the Extracellular Matrix?

The mouse mammary epithelial cell line, HC11, will not express transfected BLG and other milk proteins at high levels unless the cells are fully confluent, and have formed a basal membrane structure (Figure 6.6). This resulted in the theory that the extracellular matrix conveys a signal that maximises milk protein expression by controlling Stat5 activation levels (Zoubiane et al., 2003). GR also has been shown to have a role in maintaining differentiated mammary epithelial cells (Murtagh et al., 2004).

The formation of a basal matrix is another physical manifestation of mammary cell differentiation, this structure is essential in formation of lactationally competent cells (Roskelley et al., 1994; Streulli and Edwards, 1998). Milk protein promoters are not induced by a cocktail of dexamethasone, insulin and prolactin (D/I/P) in dividing HC11 cells, cells must be confluent and have produced a basal membrane structure (Figure 6.6). The development of the basal membrane structure in HC11 cells may mirror the differentiation process of mammary cells in late pregnancy *in vivo*. The HC11 cell line was derived from a mid pregnant mouse (Danielson et al., 1984) and does not express BLG as well as mice containing an identical transgene, this is also reflected in the strength of the HSIII (B. Whitelaw, personal communication). This could reflect that the transfected BLG is not incorporated into the optimal chromatin structure as it has not undergone development.

6.2.6 Nuclear Arrangement of Milk Protein Genes

The WAP gene is flanked by two highly expressing genes and its mammary specificity is probably at least partially controlled by a potential nuclear matrix attachment region (MAR) which has been detected in rabbit and mouse WAP at –8kb from the transcription start site (Millot et al., 2003). These sites may represent a chromatin boundary.

There is also a MAR in the 3′ flanking sequences of BLG, that appears to affect the basal level of BLG expression (Whitelaw et al., 2000). Although a shorter sequence that does not include this is also capable of directing position independent expression in transgenic mice but not to act as a locus control region (Whitelaw et al., 1992; James et al., 2000). MARs can anchor DNA sequences to the nuclear matrix, one role of this could be to localise

groups of genes, such as the milk proteins, to a common region of the nucleus. It is unknown how the locations of milk protein genes relate to each other within the nucleus, and this would be an interesting question to answer.

6.3 Stat5 Actions.

The hypothesis of this project is that a positioned nucleosome on the BLG promoter could regulate Stat5 binding to binding sites in the promoter, thus influencing BLG gene activity. It was proposed that this could occur by two separate mechanisms. Either by a nucleosome position arrangement that facilitated the Stat5 tetramerisation interaction (Figures 1.9 and 6.7), or by alternative nucleosome positions that either allowed or did not allow Stat5 to bind (Figures 1.9 and 6.7). All three Stat5 binding sites have been shown to be required for full BLG activity in the mouse mammary epithelial cell line (HC11) and in mice transgenic for BLG although even after mutation of all three Stat5 binding sites (SAA) a minimal and non-inducible expression still occurred (Burdon et al., 1994b; Burdon et al., 1994a), suggesting that hormone induced BLG expression is mediated by Stat5 action, but that other factors control the basal mammary specific expression. This leaky basal expression level may reflect the poised state of the promoter.

6.3.1 A Role for Stat5 Tetramerisation in the Regulation of BLG?

The winding of DNA round a nucleosome could bring binding sites for Stat5 close enough to allow two bound Stat5 dimers to be able to facilitate tetramerisation interactions through their N-terminal domains. A nucleosome positioned more centrally between the two outer Stat5 binding sites, such as the strong nucleosome position detected at –184bp (Figure 3.17), would favour this arrangement by bringing two bound Stat5 dimers

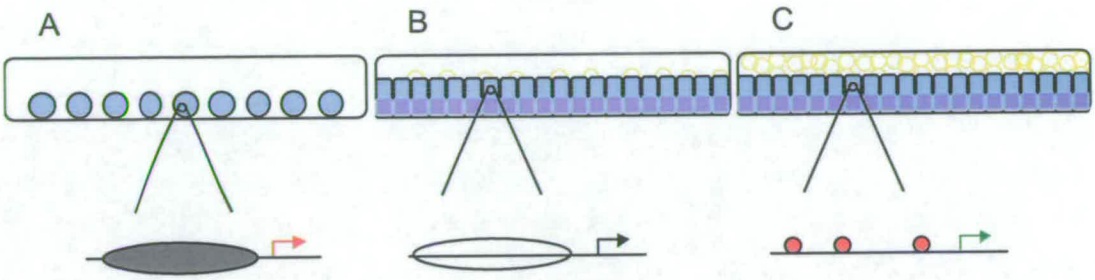


Figure 6.6. HC11 Differentiation Model.

A Schematic diagram showing how the chromatin structure in undifferentiated **(A)** dividing mammary epithelial cells in culture (top) may maintain the BLG gene in a repressed chromatin structure (grey oval, bottom). When cells reach confluence they differentiate and form a basal membrane structure (purple) **(B)** resulting in a change in the chromatin structure (open oval) that allows transcription factor binding (represented here by filled red circles) after treatment of cells with the lactogenic hormones. The subsequent activation of milk protein genes (yellow circles) may coincide with removal of the nucleosome **(C)**.

equidistant from the nucleosome (Figure 1.9, A and C). In this model Stat5 probably binds to the full consensus site StM first, this facilitates recruitment of a second dimer to the A1 and perhaps eventually also to the A3 binding site. A3 is located near to the dyad axis of the nucleosome and therefore also close to the other two Stat5 binding sites. The A1 and A3 sites are separated by 76bp, which places them together when both are in a nucleosome, such an organisation in a "supergroove" may allow two Stat5 dimers bound to these sites to interact (Edayathumangalam et al., 2004). Although Stat5 cannot bind to the A3 site *in vitro*, the nucleosome is almost certainly not present when the gene is in an active state and the A3 site will therefore be available for binding. The path of the DNA exiting the nucleosome is unknown so these interactions cannot be accurately modelled.

The ability of Stat5a to tetramerise on the BLG promoter was investigated by the addition of purified recombinant Stat5a to nucleosomes reconstituted on nAB. Complexes formed were then competed with varying amounts of a specific cold competitor DNA, an oligonucleotide to the A3 binding site. Two Stat5 dimers bound as a tetramer should form a more stable complex than two independently bound dimers. The mutant Stat5a W37A, which cannot make tetramerisation interactions but is able to bind as a dimer, was used as a control for individual dimer binding affinity. If tetramerisation interactions occur, WT Stat5a binding should be more stable than Stat5a W37A binding.

No significant stabilisation of WT Stat5a complexes compared to Stat5a W37A complexes was detected on either the N1 or the N3 nucleosome positions at high levels of competitor (Figure 5.13). However at this 200-fold molar excess competitor a Stat5 dimer bound to StM is almost completely competed (Figure 4.2). A slight stabilisation was detected for WT Stat5a over Stat5aW37A containing complexes for both the nucleosome position isomers N1 and N3 using competition with a lower range of competitors. A 20-fold but not a 10-fold molar excess of the unlabelled A3 binding site oligonucleotide produced this difference (Figure 5.14). However the difference was not observed with the same 20-fold molar excess of the Stat5 binding site oligonucleotide A3 (Figure 5.13). To be certain if tetramerisation is occurring this experiment must be repeated. It may be that the level of binding stabilisation conferred by tetramerisation on a nucleosome is very subtle and that a 20-fold molar excess competition is close to the point where the affinity of Stat5 dimers and tetramers for the probe differs. Slight variations from this or the time competition is carried out for may result in

all or both being competed. Thus it has not been proven that neither the N1 or the N3 arrangement are able to facilitate a Stat5 tetramerisation interaction nor that they cannot. The N2 position isomer has not been tested in this assay (see Section 6.2.5).

6.3.2 Recruitment of Stat5.

A second mechanism by which the nucleosome position could affect gene regulation of BLG also involves recruitment of Stat5 but does not take into account stabilisation of binding to weaker sites by tetramerisation. This model assumes that Stat5 will bind preferentially to binding sites located outwith the nucleosome. Although Stat5 has been shown to bind to sites in a nucleosome, other transcription factors that also do so have been shown to bind to free DNA with a higher affinity (Ruh et al., 2004; Li and Wrange, 1993). It can be predicted that Stat5 will bind with a similar behaviour although this remains to be tested.

The strongest Stat5 binding site, StM, is placed furthest from the nucleosome in the N3 nucleosome position. N1 also places StM a significant distance outside the nucleosome boundary (Figures 5.1 and 6.7). SP1 and GR binding sites are also situated external to the nuclesomé in N3, and lie just internal in the N1 arrangement (Figure 6.7). These transcription factors can bind to DNA arranged in a nucleosome and GR has been implicated in the disruption of a positioned nucleosome (Li and Wrange, 1993; Li et al., 1994; Perlmann and Wrange, 1988; Fletcher et al., 2000) and Chapter 5 of this thesis. Binding of Stat5 to StM may result in recruitment of CBP/p300 and GR and eventually in removal of the nucleosome so Stat5 can bind to the other two binding sites. This model could occur in any of the mapped nucleosome positions.

6.3.3 Alleviation of Repression.

Stat5 may also be able to bind equally as well to the BLG promoter in other cell types and at other promoters. Expression of β-casein has been linked to alleviation of repression, it may be that non mammary cell types contain negative regulators of Stat5 such as SOCS or the C-terminal truncated dominant negative version of Stat5 that prevent activation of the gene. A specific interaction with the prolactin receptor when the correct cocktail of lactation hormones is present in the cell may create a specifically modified Stat5 that mediates formation of a mammary specific complex over the promoter.

Stat5 is constitutively phosphorylated on serine 779 in mammary tissue (Beuvink et al., 2000). This modification appears to prevent Stat5 binding to DNA and reduces the expression level of a reporter gene driven by the β-casein promoter but is overcome by co-stimulation of cells with dexamethasone (a GR agonist) and prolactin. The mechanism of phospho-serine inhibition is unknown. The function of this modification could be to repress milk production during late pregnancy and mammary gland development, when the poised chromatin structure may have developed, until a cooperative interaction with GR overcomes the phospho-serine inhibition. By this mechanism, although high levels of Stat5 are present from late pregnancy, milk protein genes are not activated until parturition (Yamashita et al., 2001).

6.3.4 Does Stat5 Position Nucleosomes.

If there is not a positioned nucleosome in the active gene, there is still a strong positioning signal present in the DNA, such that the nucleosome must

be actively removed and prevented from binding here. The *in vitro* map of nucleosome positions over the BLG gene matched the *in* vivo map in liver, suggesting that this is the default state of the gene (Boa, 1999). The above models (Figure 6.7) provide a mechanism by which Stat5 binding facilitates disruption of this nucleosome. An alternative model proposes that Stat5 binding resuts in the poised chromatin structure.

Stat5 is required for mammary gland development. At some point during this development I have proposed that a chromatin structure over the BLG promoter is formed that renders the gene in a poised state, ready to be activated. This has yet to be tested by looking at the chromatin structure *in vivo* over the promoter, but HSIV and V suggest a different chromatin structure is present over the non expressed BLG gene in mammary gland compared to that observed in a tissue that BLG is never expressed in. These hypersensitive sites lie in the transcribed region of the gene, not over the proximal promoter where a Stat5 dependent chromatin structure is predicted to take place, however a chromatin structure could be nucleated at the proximal promoter and spread along the gene in a similar way to heterochromatin spreading or spread to the proximal promoter from a different nucleation point.

One argument against Stat5 creating a poised chromatin structure, is that no displacement of nucleosomes was detected after Stat5 binding, however it is likely that if this occurs it is mediated by interactions with other proteins recruited to the promoter by Stat5, for example GR. Bandshifts with lactating mammary nuclear extracts or extracts from cultured cells which can be treated with various stimulus may reveal more about the role of Stat5 and

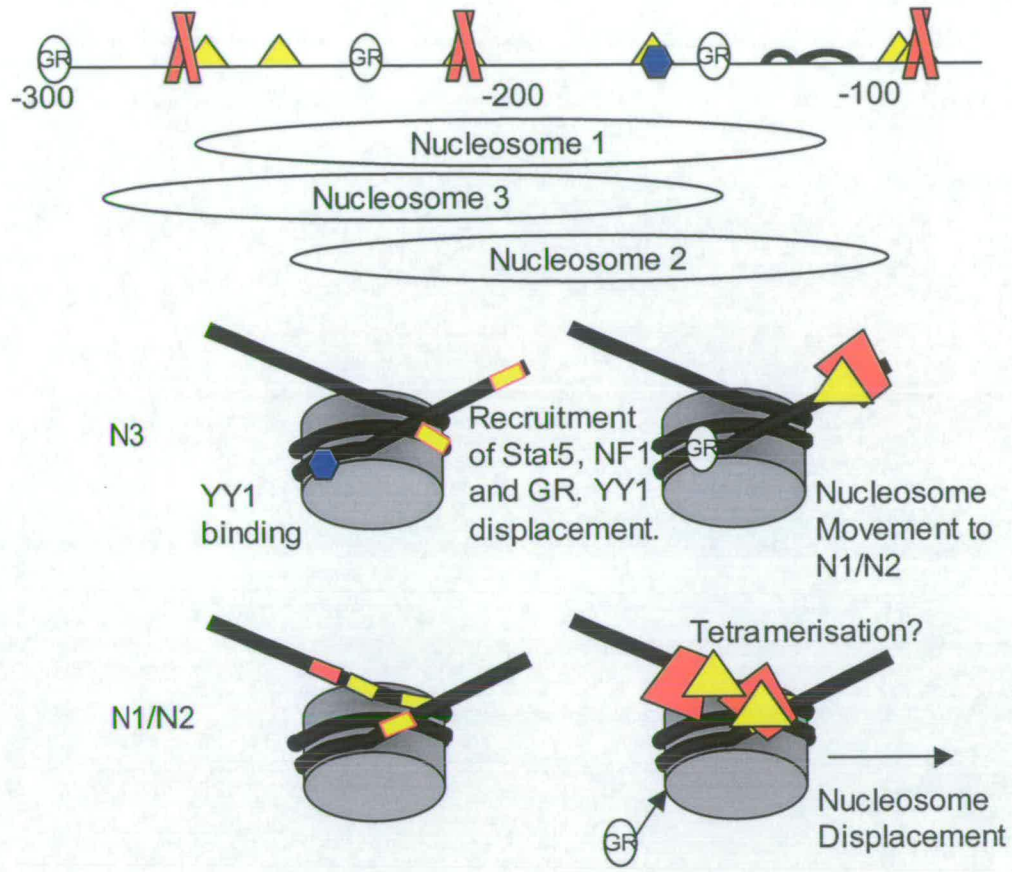


Figure 6.7 Transcription Factors Binding to Nucleosomes.

Schematic diagrams depicting transcription factors binding to nAB in nucleosome positions N3 and N1/N2. Transcription factors are represented by the same shapes as in Figure 6.4. In the N3 position, Stat5 is recruited to the StM binding site and GR can also bind, Stat5 may not be able to bind to the A1 and A3 binding sites. YY1 may also be able to bind in N3. In N1/N2 Stat5 can bind to both StM and A1 which may potentially allow tetramerisation. SP1 and GR binding may induce nucleosome displacement and allow NF1 binding.

other factors in nucleosome movement.

6.3.5 Functional Significance of Tetramerisation.

Work is in progress to compare the effects of organisation of the BLG promoter into chromatin on expression level of the gene. The BLG promoter is either transiently or stably transfected into CHO-K3 cells. Transient transfections will be organised into a form of chromatin structure, but stable integrants may have a more organised chromatin structure formed over them, which may be organised into an appropriate structure by the signals in the DNA. This approach however does not mimic mammary gland differentiation. If any difference is observed between the stably and transiently transfected cells then this experiment could be taken a stage further. The effects of overexpression of the tetramerisation mutant and WT Stat5a and 5b will be compared in both types of transfected cell with and without D/I/P stimulation (Neil et al., 2004).

6.4 The N2/ NX Positions.

The N2 position isomer exhibits an interesting behaviour. N2 displays a tendency to migrate as the isomer NX. It does so with minor changes in the lengths of tail DNA that lie external to the nucleosome (Figures 3.6 and 3.8). This led to the conclusion that the NX position may already exist within the N2 band, as a different position that migrates through a non-denaturing polyacrylamide gel at the same rate as N2. An isomer migrating in the NX band also appears when the A3 Stat5 binding site is mutated to A3m, by exchange of an A to a T and insertion of a T (Figures 5.2 and Figure 4.1). It is unlikely that this single insertion would be sufficient to significantly alter the migration of a positioned nucleosome but it may alter the nucleosome position. The change lies within 30 bp of the dyad axis, this is the DNA that

makes contact with the $(H3H4)_2$ tetramer, which is responsible for positioning the nucleosome. It is possible that this 1bp insertion is sufficient to cause an entirely new nucleosome position on nAB. Groups searching for the signal in the DNA that drives nucleosomes positioning *in vitro* could in the future exploit single base changes like this.

Another feature of the N2 isomer is that approximately 30bp of DNA upstream of the nucleosome core interacts with the nucleosome. This interaction is sufficient to cause a non sequence-specific pause in an ExoIII digest that does not remain at higher ExoIII concentrations suggesting that it is not a true nucleosome pause site (Figure 3.11). The N1 site lies upstream of the N2 site and the DNA interacting with the nucleosome is the sequence that separates the two position isomers. This interaction may represent a link between N2 and N1. Alternatively, the extra DNA covered by this structure places the two outer Stat5 binding sites exactly at the nucleosome boundary. The centre of this extended nucleosome would lie at –184. The significance of this arrangement is not yet clear.

In the restriction enzyme accessibility assays of the nucleosome position isomers in Chapter 5, N2 also exhibited a strange behaviour. A *Rsa*I site lying 22bp upstream of the nucleosome boundary was not cleaved by *Rsa*I, but *Hin*fI cleaved at a site 30bp towards the nucleosome dyad (Figure 5.11), this *Hin*fI site is within the N2 nucleosome (Figure 5.12). *Hin*fI cuts within N1 at both sites better than it does N2, this may reflect a stronger interaction of the DNA with the nucleosome core in N2 than in N1, perhaps facilitated by the extra DNA associated with the nucleosome. This may reflect a

property of the N2 nucleosome *in vitro* where transcription factor access is impeded.

In Stat5a bandshifts of unfractionated reconstitutes, Stat5 preferentially bound to the N2 nucleosome position isomer over either the N1 or the N3 isomers (Figure 5.2). The additional 30bp associated with the nucleosome brings both the Stat5 binding sites together right at the boundary of the nucleosome associated DNA. Stat5 has been demonstrated to be able to bind to DNA at the periphery of a nucleosome, it is therefore likely that both these Stat5 binding sites can be bound in this extended nucleosome, as long as binding of one does not impede the binding of the other. As the N2 isomer also exhibits a tendency to reposition to the isomer NX, and thus the location of this nucleosome on nAB could not be established beyond doubt, competition assays to determine the affinity of WT and W37A Stat5a for these positions were not carried out. Given a more accurate location of N2 and NX, perhaps by hydroxyl radical mapping, it would be of interest to study the relative affinity of N2 and NX in the competition assay.

6.5 Which Stat5 Sites are Bound *In Vivo.*

Binding to all three Stat5 binding sites is essential for full gene activity in mammary epithelial cells. Conventional chromatin immunoprecipitation is unable to differentiate between the three binding sites, as all would be cross linked in one large complex by the nucleosome. A technique such as incorporation of a DNA cleavage or crosslinking reagent into the Stat5 molecule may be able to determine exactly which Stat5 binding sites are occupied, and at which stage of mammary gland development. An alternative strategy could be to carry out an *in vivo* footprint of the region,

involving digest of the genomic DNA, perhaps with ExoIII or DNaseI and making use of LM-PCR to detect protected areas over the Stat5 binding sites.

6.6 Complex Regulation of BLG.

Work in this thesis has shown that Stat5 can bind to sites at the boundary of a a nucleosome, but not to sites further than 40bp from the boundary. Nucleosome position isomers have been mapped and a specific feature of the N2 nucleosome isomer described. A role of Stat5 tetramerisation in the binding of Stat5a to DNA within a nucleosome was investigated, although this study was not conclusive.

From these results I have proposed several models for the mammary specific activation of the BLG gene. Many of these will be relevant to other milk proteins. The nature of the mammary specific signal for BLG expression is yet to be discovered and many avenues of investigation remain open.

Reference List

Adkins,N.L., Watts,M., and Georgel,P.T. (2004). To the 30-nm chromatin fiber and beyond. Biochimica et Biophysica Acta *1677*, 12-23.

Aigueperse,C., Val,P., Pacot,C., Darne,C., Lalli,E., Sassone-Corsi,P., Veyssiere,G., Jean,C., and Martinez,A. (2001). SF-1 (Steroidogenic Factor-1), C/EBPβ (CCAAT/Enhancer Binding Protein), and Ubiquitous Transcription factors NF-1 (Nuclear Factor 1) and Sp1 (Selective Promoter factor 1) Are Required for Regulation of the Mouse Aldose Reductase-like Gene (AKR1B7) Expression in Adrenocortical Cells. Molecular Endocrinology *15*, 93-111.

Alexander,M., Heppel,L.A., and Hurwitz,J. (1961). The Purification and Properties of Micrococcal Nuclease. J. Biol. Chem *236*, 3014-3019.

Anderson,J.D., Thåström,A., and Widom,J. (2002). Spontaneous Access of proteins to buried Nucleosomal DNA Target Sites Occurs via a Mechanism That is Distinct from Nucleosome Translocation. Mol Cell Biol *22*, 7147-7157.

Archibald,A.L., McClenaghan,M., Hornsey,V., Simons,J.P., and Clark,A.J. (1990). High-level expression of biologically active human alpha 1-antitrypsin in the milk of transgenic mice. Proc Natl Acad Sci USA *87*, 5178-5182.

Barahmand-Pour,F., Meinke,A., Groner,B., and Decker,T. (1998). Jak2-Stat5 interactions analyzed in yeast. J. Biol. Chem *273*, 12567-12575.

Beato,M. and Eisfeld,K. (1997). Transcription Factor Access to Chromatin. Nucleic Acids Research *25*, 3559-3563.

Becker,P.B. and Horz,W. (2002). ATP-Dependent Nucleosome Remodeling. Annu Rev Biochem *71*, 3247-273.

Becker,S., Groner,B., and Muller,C.W. (1998). Three-dimensional structure of the Stat3beta homodimer bound to DNA. Nature *394*, 145-151.

Bendich,A.J. and Drlica,K. (2000). Prokaryotic and eukaryotic chromosomes: what's the difference? BioEssays *22*, 481-486.

Bergad,P.L., H,M.S., Towle,H.C., Schwarzengberg,S.J., and Berry,S.A. (1995). Growth hormone induction of hepatic serine protease inhibitor 2.1 transcription is mediated by a Stat5-related factor binding synergystically to two γ-activated sites. J. Biol. Chem *270*, 24903-24910.

Beuvink,I., Hess,D., Flotow,H., Hofsteenge,J., Groner,B., and Hynes,N.E. (2000). Stat5a Serine Phosphorylation. J. Biol. Chem *275*, 10247-10255.

Biener,E., Martin,C., Daniel,N., Frank,S.J., Centonze,V.E., Herman,B., Djiane,J., and Gertler,A. (2003). Ovine PLacental Lactogen-Induced Heterodimerization of Ovine Growth Hormone and Prolactin Receptors in Living Cells Is Demonstrated by FIluorescence Resonance Energy Transfer Microscopy and Leads to Prolonged Phosphorylation of Signal Transduceer and Activator of Transcription (STAT)1 and STAT3. Endocrinology *144*, 3532-3540.

Boa, S. A. Nucleosomal Organisation over the Ovine β–Lactoglobulin Gene. 1999. University of Edinburgh.
Ref Type: Thesis/Dissertation

Boeger,H., Griesenbeck,J., Strattan,J.S., and Kornberg,R.D. (2003). Nucleosomes Unfold Completely at a Transcriptionally Active Promoter. Molecular Cell *11*, 1587-1598.

Bowen,N.J. and Jordan,I.K. (2002). Transposable elements and the evolution of eukaryotic complexity. Current Issues in Molecular Biology *4*, 65-76.

Brelje,T.C., Svensson,A.M., Stout,L.E., Bhagroo,N.V., and Sorenson,R.L. (2002). An immunohistochemical approach to monitor the prolactin-induced activation of the JAK2/STAT5 pathway in pancreatic islets of Langerhans. J Histochem Cytochem *50*, 365-383.

Bruno,M., Flaus,A., Stockdale,C., Rencurel,C., Ferreira,H., and Owen-Hughes,T. (2003). Histone H2A/H2B Dimer Exchange by ATP-Dependent Chromatin Remodeling Activities. Molecular Cell *12*, 1599-1606.

Buckle,R., Balmer,M., Yenidunya,A., and Allan,J. (1991). The promoter and enhancer of the inactive chicken beta-globin gene contains precisely positioned nucleosomes. Nucleic Acids Research *19*, 1219-1226.

Burdon,T.G., Demmer,J., Clark,A.J., and Watson,C.J. (1994a). The Mammary Factor MPBF is a Prolactin-Induced Transcriptional Regulator which Binds to STAT Factor Recognition sites. Febs Lett *350*, 177-182.

Burdon,T.G., Maitland,K.A., Clark,A.J., Wallace,R., and Watson,C.J. (1994b). Regulation of the Sheep Beta-Lactoglobulin Gene by Lactogenic Hormones is Mediated by a Transcription Factor that Binds an Interferon-Gamma Activation Site-Related Element. Molecular Endocrinology *8*, 1528-1536.

Buttinelli,M., Di Mauro,E., and Negri,R. (1993). Multiple nucleosome positioning with unique rotational setting for the *saccharomyces cerevisiae* 5S rRNA gene *in vitro* and *in vivo*. Proc Natl Acad Sci USA *90*, 9319.

Carruthers,L.M., Bednar,J., Woodcock,C.L., and Hansen,J.C. (1998). Linker Histones Stabilize the Intrinsic Salt-Dependent Folding of Nucleosomal Arrays: Mechanistic ramifications for Higher-Order Chromatin Folding. Biochemistry *37*, 14776-14787.

Cella,N., Groner,B., and Hynes,N.E. (1998). Characterization of Stat5a and Stat5b Homodimers and heterodimers and their association with the Glucocorticoid Receptor in Mammary Cells. Mol Cell Biol *18*, 1783-1792.

Chapman,R.S., Lourenco,P.C., Tonner,E., Flint,D.J., Selbert,S., Takeda,K., Akira,S., Clarke,A.R., and Watson,C.J. (1999). Suppression of epithelial apoptosis and delayed mammary gland involution in mice with a conditional knockout of Stat3. Genes and Development *13*, 2604-2616.

Chen,X., bhandari,R., Vinkemeier,U., van den Akker,F., Darnell,J.E., Jr., and Kuriyan,J. (2002). A reinterpretation of the dimerization interface of the N-terminal Domains of STATs. Protein Science *12*, 361-365.

Chen,X., Vinkemeier,U., Zhao,Y., Jeruzalmi,D., Darnell,J.E., and Kuriyan,J. (1998). Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. Cell *93*, 827-839.

Cheung,E., Schwabish,M.A., and Kraus,W.L. (2003). Chromatin exposes intrinsic differences in the transcriptional activities of estrogen receptors α and β. EMBO J *22*, 600-611.

Copeland,N.G., Gilbert,D.J., Schindler,C., Zhong,Z., Wen,Z., Darnell,J.E.J., Mui,A.L.F., Miyajima,A., uelle,F.W., hle,J.N., and enkins,N.A. (1995). Distribution of the Mammalian *Stat* Gene Family in Mouse Chromosomes. Genomics *29*, 225-228.

Cordingley,M.G., Rigel,A.T., and Hager,G.L. (1987). Steroid-Dependent Interaction of Transcription Factors with the Inducible Promoter of Mouse Mammary Tumour Virus *In Vivo*. Cell *48*, 261-271.

Corlata,I., Ren,S., Zhang,Y., Gehan,E., Baljit,S., and Furth,P.A. (2004). Stat5a in Tyrosine Phosphorylated and Nuclear Localised in a High Proportion of Breat Cancers. International Journal of Cancer *108*, 665-671.

Cosma,M.P. (2002). Ordered Recruitment: Gene-Specific mechanism of Transcription Activation. Molecular Cell *10*, 227-236.

Cousins,D.J., Islam,S.A., Sanderson,M.R., Proykoya,Y.G., Crane-Robinson,C., and Staynov,D.Z. (2004). Redefinition of the Cleavage Sites of DNaseI on the Nucleosome Core particle. J. Mol Biol *335*, 1199-1211.

Cuatrecasas,P., Fuchs,S., and Anfinsen,C.B. (1967). Catalytic Properties and Specificity of the Extracelllar Nuclease of *Staphylococcus aureus*. J. Biol. Chem *242*, 1541-1547.

Danielson,K.G., Oborn,C.J., Durban,E.M., Butel,J.S., and Medina,D. (1984). Epithelial mouse mammary cell line exhibiting normal morphogenesis in vivo and functional differentiation in vitro. Proc Natl Acad Sci USA *81*, 3756-3760.

Darnell,J.E. (1997). STATs and gene regulation. Science *277*, 1630-1635.

Darnell,J.E.J., Kerr,I.M., and Stark,G.R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. Science *264*, 1415-1421.

Davey,C., Fraser,R., Smolle,M.W., Simmen,M.W., and Allan,J. (2003). Nucleosome Positioning Signals in the DNA Sequence of the Human and Mouse H19 Imprinting Control regions. J. Mol Biol *325*, 873-887.

Davey,C., Pennings,S., Meersseman,G., Wess,T.J., and Allan,J. (1995). Periodicity of strong nucleosome positioning sites around the chicken adult beta-globin gene may encode regularly spaced chromatin. Proc Natl Acad Sci USA *92*, 11210-11214.

Demmer,J., Burdon,T.G., Djiane,J., Watson,C.J., and Clark,A.J. (1995). The proximal milk protein binding factor binding site is required for the prolactin responsiveness of the sheep beta-lactoglobulin promoter in Chinese hamster ovary cells. Mol Cell Endocrinol *107*, 113-121.

Dignam,J.D., Lebovitz,R.M., and Roeder,R.G. (1983). Accurate transcription Initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Research *11*, 1475-1489.

Dobie,K., Mehtali,M., McClenaghan,M., and Lathe,R. (1997). Variegated gene expression in mice. Trends in Genetics *13*, 127-130.

Dong,F., Hansen,J.C., and van Holde,K.E. (1990). DNA and protein determinants of nucleosome positoning on sea urchin 5S rRNA gene sequences *in vitro*. Proc Natl Acad Sci USA *87*, 5724-5728.

Dong,F. and van Holde,K.E. (1991). Nucleosome positioning is determined by the $(H3\text{-}H4)_2$ tetramer. Proc Natl Acad Sci USA *88*, 10596-19600.

Edayathumangalam,R.S., Weyermann,P., Gottesfeld,J.M., Dervan,P.B., and Luger,K. (2004). Molecular recognition of the nucleosomal "supergroove". Proc Natl Acad Sci USA *101*, 6864-6869.

Eisfeld,K., Candau,R., Truss,M., and Beato,M. (1997). Binding of NF-1 to the MMTV promoter in nucleosomes: influence of rotational phasing, translational positioning and histone H1. Nucleic Acids Research *25*, 3733-3742.

Felsenfeld,G., Boyes,J., Clark,D., and Studitsky,V. (1996). Chromatin structure and gene expression. Proc Natl Acad Sci USA *93*, 9384-9388.

Festenstein,R. and Kioussis,D. (2000). Locus control regions and epigenetic chromatin modifiers. Curr Op Gen Dev *10*, 199-203.

FitzGerald,P.C. and Simpson,R.T. (1985). Effects of Sequence Alterations in a DNA Segment Contining the 5 S RNA Gene from *Lytechinus variegatus* on Positioning of a Nucleosome Core Particle *in Vitro*. J. Biol. Chem 15318-15324.

Flaus,A., Luger,K., Tan,S., and Richmond,T.J. (1996). Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals. Proc Natl Acad Sci USA *93*, 1375.

Flavin,M., Cappabianca,L., Kress,C., Thomassin,H., and Grange,T. (2004). Nature of the Accessible Chromatin at a Glucocorticoid-Responsive Enhancer. Mol Cell Biol *24*, 7891-7901.

Fletcher,T.M., Ryu,B.W., Baumann,C.T., Warren,B.S., Fragoso,G., John,S., and Hager,G.L. (2000). Structure and dynamic properties of a glucocorticoid receptor- induced chromatin transition. Mol Cell Biol *20*, 6466-6475.

Gencheva, M. and Allan, J. 2005.
Ref Type: Unpublished Work

Gewinner,C., Hart,G., Zachara,N., Cole,R., Beisenherz-Huss,C., and Groner,B. (2004). The Coactivator of Transcription CREB-binding protein Interacts Preferentially with the Glycosylated Form of Stat5. J. Biol. Chem *279*, 3563-3572.

Gottesfeld,J.M. (1987). DNA sequence-directed nuceosome reconstitution on 5S RNA genes of Xenopus laevis. Mol Cell Biol *7*, 1612-1622.

Grewal,S.I.S. and Elgin,S.C. (2002). Heterochromatin: new possibilities for the inheritance of structure. Curr Op Gen Dev *12*, 178-187.

Hansen,J.C. and van Holde,K.E. (1991). The mechanism of Nucleosome Assembly onto Oligomers of the Sea Urchin 5 S DNA Positioning Sequence. J. Biol. Chem *266*, 4276-4282.

Harris,S., McLenaghan,M., Simons,J.P., Ali,S., and Clark,A.J. (1991). Developmental regulation of the sheep beta-lactoglobulin gene in the mammary gland of transgenic mice. Developmental Genetics *12*, 299-307.

Heim,M.H. (1999). The Jak-STAT pathway: Cytokine signalling from the receptor to the nucleus. Journal Of Receptor And Signal Transduction Research *19*, 75-120.

Heitz,E. (1928). Das heterochromatin der Moose. Jehrb Wiss Botanik *69*, 762-818.

Hickey,D.A. (1992). Evolutionary dynamics of transposable elements in prokaryotes and eukaryotes. Genetica *86*, 269-274.

Hoey,T. and Schindler,U. (1998). STAT structure and function in signaling. Curr Op Gen Dev *8*, 582-587.

Hogan,M.E., Rooney,T.F., and Austin,R.H. (1987). Evidence for kinks in DNA folding in the nucleosome. Nature *328*, 554-557.

Horn,P.J. and Peterson,C.L. (2002). Chromatin Higher Order Folding, Wrapping up Transcription. Science *297*, 1824-1827.

Horvath,C.M. and Darnell,J.E.J. (1997). The state of the STATs: recent developments in the study of Signal Transduction to the nucleus. Current Opinion in Cell Biology *9*, 233-239.

Horz,W. and Altenburger,W. (1981). Sequence specific cleavage of DNA by micrococcal nuclease. Nucleic Acids Research *9*, 2643-2658.

Hou,X.S., Melnick,M.B., and Perrimon,N. (1996). Marelle acts downstream of the Drosophila HOP/JAK kinase and encodes a protein similar to the mammalian STATs. Cell *84*, 411-419.

Iavnilovitch,E., Groner,B., and Barash.I (2002). Overexpression and Forced Activation of Stat5 in mammary Gland of Transgenic Mice Promotes Cellular Proliferation, Enhances Differentiation and Delays Postlactational Apoptosis. Molecular Cancer Research *1*, 32-47.

Ihle,J.N. (1996). STATs: Signal Transducers and Activators of Transcription. Cell *84*, 331-334.

Jackson,j.R. and Benyajati,C. (1993). DNA-Histone Interactions are Sufficient to Position a Single Nucleosome Juxtaposing *Drosophila Adh* Adult Enhancer and Distal Promoter. Nucleic Acids Research *21*, 957-967.

James,R.M., Neil,C., Webster,J., Roos,S., Clark,A.J., and Whitelaw,C.B.A. (2000). Multiple Copies of β-Lactoglobulin Promoter Do Not Function as LCR. Biochemical and Biophysical Research Communications *272*, 284-289.

Jenuwein,T. and Allis,C.D. (2001). Translating the Histone Code. Science *293*, 1074-1080.

John,S., Robbins,C.M., and Leonard,W.J. (1996). An IL-2 response element in the human IL-2 receptor α chain receptor is a composite element that binds Stat5, Elf-1 HMG-I(Y) and a GATA family protein. EMBO J *15*, 5627-5635.

John,S., Vinkemeier,U., Soldaini,E., Darnell,J.E., and Leonard,W.J. (1999). The significance of tetramerization in promoter recruitment by Stat5. Mol Cell Biol *19*, 1910-1918.

Jolivet,G., Devinoy,E., Fontaine,M.L., and Houdebine,L.-M. (1992). Structure of the Gene Encoding Rabbit Alpha s1-Casein. Gene *113*, 257-62.

Jolivet,G., L'Hotte,C., Pierre,S., Tourkine,N., and Houdebine L-M. (1996). A MGF/STAT5 binding site is necessary in the distal enhancer for higher prolactin induction of transfected rabbit αS1-*casein-CAT* gene transcription. Febs Lett *389*, 257-262.

Juergens,W.G., Stockdale,F.E., Topper,Y.J., and Elias,J.J. (1965). Hormone-dependent differentiation of mammary gland *in vitro*. Proc Natl Acad Sci USA *54*, 629-634.

Kabotyanski,E.B. and Rosen,J.M. (2003). Signal Transduction Pathways Regulated by Prolactin and Src Result in Different Conformations of Activated Stat5b. J. Biol. Chem *278*, 17218-12227.

Kawata,T., Shevchenko,A., Fukuzawa,M., Jermyn,K.A., Totty,N.F., Zhukovskaya,N.V., Sterling,A.E., Mann,M., and Williams,J.G. (1997). SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in dictyostelium. Cell *89*, 909-16.

Kazansky,A.V., Kabotyanski,E.B., Wyszomierski,S.L., Mancini,M.A., and Rosen,J.M. (1999). Differential Effects of prolactin and src/abl kinases on the Nuclear Translocation of STAT5B and STAT5A. J. Biol. Chem *274*, 22484-22492.

Kefalas,P., Gray,F.C., and Allan,J. (1988). Precise nucleosome positioning in the promoter of the chicken beta A globin gene. Nucleic Acids Research *16*, 501-517.

Kelly,P.A., Bachelot,A., Kedzia,C., Hennighausen,L., Ormandy,C.J., Kopchick,J.J., and Binart,N. (2002). The Role of Prolactin and Growth Hormone in Mammary Gland Development. Molecular And Cellular Endocrinology *197*, 127-131.

Kim,A. and Dean,A. (2004). Developmental Stage Differences in Chromatin Subdomains of the β-Globin Locus. Proc Natl Acad Sci USA *101*, 7028-7033.

Kontopodis,G., Holt,C., and Sawyer,L. (2002). The Ligand-binding Site of Bovine β-Lactoglobulin: Evidence for a function? J. Mol Biol *318*, 1043-1055.

Lee,C.-K., Shibata,Y., Rao,B., Strahl,B.D., and Lieb,J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome wide. Nature Genetics 1-6.

Lee,C.S. and Oka,T. (1992). A Pregnancy-Specific Mammary Nuclear factor Involved in thr Repression of the Mouse β-casein Gene Transcription by Progesterone. J. Biol. Chem *267*, 5797-5801.

Lee,K.-F., Atiee,S.H., and Rosen,J.M. (1989). Differential regulation of Rat β-casein-Chloramphenicol Acetyltransferase Fusion gene Expression in Transgenic Mice. Mol Cell Biol *9*, 560-565.

Li,B., Adams,C.C., and Workman,J.L. (1994). Nucleosome Binding by the Constitutive Transcription Factor Sp1. J. Biol. Chem *10*, 7756-7763.

Li,Q. and Wrange,O. (1993). Translational positioning of a nucleosomal glucocorticoid response element modulates a glucocorticoid receptor affinity. Genes and Development *7*, 2471-2482.

Li,S. and Rosen,J.M. (1995). Nuclear factor I and mammary gland factor (STAT5) play a critical role in regulating rat whey acidic protein gene expression in transgenic mice. Mol Cell Biol *15*, 2063-2070.

Lin,J.-X., Mietz,J., Modi,W.S., John,S., and Leonard,W.J. (1996). Cloning of Human Stat5B. J. Biol. Chem *271*, 10738-10744.

Linxweiler,W. and Horz,W. (1982). Sequence Specificity of exonuclease III from *E.coli*. Nucleic Acids Research *10*, 4845-4859.

Linxweiler,W. and Horz,W. (1984). Reconstitution of mononucleosomes: characterisation of distinct particles that differ in the position of the histone core. Nucleic Acids Research *12*, 9395-9413.

Linxweiler,W. and Horz,W. (1985). Reconstitution Experiments show That Sequence-Specific Histone-DNA Interactions Are the Basis for Nucleosome Phasing on Mouse satellite DNA. Cell *42*, 281-290.

Litterst,C.M., Kliem,D.M., Marilley,D., and Pfitzner,E. (2003). NCoA-1/SRC-1 Is an Essential Coactivator of STAT5 That Binds to the FDL Motif in the α-Helical Region of the STAT5 Transactivation Domain. J. Biol. Chem *278*, 45340-45351.

Liu,X., Robinson,G.W., Gouilleux,F., Groner,B., and Hennighausen,L. (1995). Cloning and Expression of Stat5 and an Additional Homologue (Stat5b) Involved in Prolactin Signal Transduction in Mouse Mammary Tissue. Proc Natl Acad Sci USA *92*, 8831-8835.

Liu,X., Robinson,G.W., Wagner,K.-U., Garrett,L., Wynshaw-Boris,A., and Hennighausen,L. (1997). Stat5a is mandatory for adult mammary gland development and lactogenesis. Genes and Development *11*, 179-186.

Lohr,D. (1997). Nucleosome Transactions on the Promoters of the Yeast *GAL* and *PHO* Genes. J. Biol. Chem *272*, 26795-26798.

Long,W., Wagner,K.-U., Lloyd,K.C., Binart,N., Shillingford,J.M., Hennighausen,L., and Jones,F.E. (2003). Impaired differentiation and lactational failure of *Erbb4*-deficient mammary glands identfy ERBB4 as an obligate mediator of STAT5. Development and Disease *130*, 5257-5268.

Lubon,H. and Hennighausen,L. (1987). Nuclear proteins from lactating mammary glands bind to the promoter of a milk protein gene. Nucleic Acids Research *11*, 2103-2121.

Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F., and Richmond,T.J. (1997). Crystal Structure of the Nucleosome Core Particle at 2.8A resolution. Nature *389*, 251-260.

Luo,G. and Yu-Lee,L.Y. (2003). Transcription Inhibition by Stat5. J. Biol. Chem *43*, 26841-26849.

Lusser,A. (2002). Acetylated, methylated, remodeled: chromatin states for gene regulation. Current Opinion in Plant Biology *5*, 437-443.

Maeda,S., Kawai,T., Obinata,M., Fujiwara,H., Horiuchi,T., Saeki,Y., Sato,Y., and Furusawa,M. (1985). Production of human α-interferon in silkworm using a baculovirus vector. Nature *315*, 592-594.

Martinez-Campa,C., Politis,P., Moreau,J.-L., Kent,N., Goodall,J., Mellor,J., and Goding,C.R. (2004). Precise nucleosome Positioning and the TATA Box Dictate Requirements for the Histone Tail and the Bromodomain Factor Bdf1. Molecular Cell *15*, 69-81.

Martino,A., Holmes,J.H.I., Lord,J.D., Moon,J.J., and Nelson,B.H. (2001). Stat5 and SP1 Regulate Transcription of the Cyclin D2 Gene in Response to IL-2. The Journal of Immunology *166*, 1723-1729.

Mayr,S., Welte,T., Windegger,M., Lechner,J., Heinrich,P.C., Horn,F., and Doppler ,W. (1998). Selective coupling of STAT factors to the mouse prolactin receptor. European Journal of Biochemistry *258*, 784-793.

Meersseman,G., Pennings,S., and Bradbury,E.M. (1991). Chromatosome Positioning on Assembled Long Chromatin. Linker Histones Affect Nucleosome Placement on 5S rDNA. J. Mol Biol *220*, 89-100.

Meersseman,G., Pennings,S., and Bradbury,E.M. (1992). Mobile nucleosomes - a general behaviour. EMBO J *11*, 2951-2959.

Meier,V.S. and Groner,B. (1994). The Nuclear Factor YY1 PParticipates In Repression of the β-Casein Gene promoter in Mammary Epithelial Cells and is Counteracted by Mammary Gland Factor during Lactogenic Hormone Induction. Mol Cell Biol *14*, 128-137.

Meinke,A., Barahmand-Pour,F., Wöhrl,S., Stoiber,D., and Decker,T. (1996). Activation of Different Stat5 Isoforms Contributes to Cell-Type-Restricted Signaling in Response to Interferons. Mol Cell Biol *16*, 6937-6944.

Meyer,W.K.H., Reichenbach,P., Schindler,U., Soldaini,E., and Nabholz,M. (1997). Interaction of STAT5 dimers on two low affinity binding sites mediates interleukin 2 (IL-2) stimulation of IL-2 receptor alpha gene transcription. J. Biol. Chem *272*, 31821-31828.

Millot,B., Fontaine,M.L., Thepot,D., and Devinoy,E. (2001). A distal region, hypersensitive to DNase I, plays a key role in regulating rabbit whey acidic protein gene expression. Biochem J *359*, 3-65.

Millot,B., Montoliu,L., Fontaine,M.L., Mata,T., and Devinoy,E. (2003). Hormone-induced modifications of the chromatin structure surrounding upstream regulatory regions conserved between the mouse and rabbit whey acidic protein genes. Biochem J *372*, 41-52.

Miyamoto,C., Smith,G.E., Farrell-Towt,J., Chizzonite,R., Summers,M.D., and Ju,G. (1985). Production of Human *c-myc* Protein in Insect Cells Infected with a Baculovirus Expression Vector. Mol Cell Biol *5*, 2860-2865.

Miyoshi,K., Shillingford,J.M., Smith,G.H., Grimm,S.L., Wagner,K.U., Oka,T., Rosen,J.M., Robinson,G.W., and Hennighausen,L. (2001). Signal transducer and activator of transcription (Stat) 5 controls the proliferation and differentiation of mammary alveolar epithelium. J Cell Biol *155*, 531-542.

Moriggl,R., Gouilleux-Gruart,V., Jähne,R., Berchtold,S., Gartmann,C., Liu,X., Hennighausen,L., Sotiropoulos,A., Groner,B., and Gouilleux,F. (1996). Deletion of the Carboxyl-Terminal Transactivation Domain of MGF-Stat5 Results in Sustained DNA Binding and a Dominant Negative Phenotype. Mol Cell Biol *16*, 5691-5700.

Mukhopadhyay,S.S., Wyszomierski,S.L., Gronostajski,R.M., and Rosen,J.M. (2001). Differential interactions of specific nuclear factor I isoforms with the glucocorticoid receptor and STAT5 in the cooperative regulation of WAP gene transcription. Mol Cell Biol *21*, 6859-6869.

Murtagh,J., McArdle,E., Gilligan,E., Thornton,L., Furlong,F., and Martin,F. (2004). Organization of mammary epithelial cells into 3D acinar structures requires glucocorticoid and JNK signaling. J Cell Biol *166*, 133-143.

Neil, C., Little, G. H., and Whitelaw, C. B. 2004.
Ref Type: Unpublished Work

Noll,M. and Kornberg,R.D. (1977). Action of micrococcal Nuclease on Chromatin and the location of Histone H1. J. Mol Biol *109*, 393-404.

Ooi,G.T., Hurst,K.R., Poy,M.N., Rechler,M.M., and Boisclair,Y.R. (1998). Binding of a Single Element resembling a γ–Interferon-Activated Sequence Mediates the Growth Hormone Induction of the Mouse Acid-Labile Subunit Promoter in Liver Cells. Molecular Endocrinology *12*, 675-687.

Ostland Farrants,A.K., Blomquist,P., Kwon,H., and Wrange,O. (1997). Glucocorticoid receptor-glucocorticoid response element binding stimulates nucleosome disruption by the SWI/SNF complex. Mol Cell Biol *17*, 895-905.

Ota,N., Brett,T.J., Murphy,T.L., Fremont,D.H., and Murphy,K.M. (2004). N-domain-dependent nonphosphorylated STAT4 dimers required for cytokine-driven activation. Nature immunology *5*, 208-215.

Panetta,G., Buttinelli,M., Flaus,A., Richmond,T.J., and Rhodes,D. (1998). Differential Nucleosome Positioning on *Xenopus* Oocyte and Somatic 5 S RNA Genes Determines both TFIIIA and H1 Binding: A Mechanism for Selective H1 Repression. J. Mol Biol *282*, 683-697.

Park,S.H., Yamashita,H., Rui,H., and Waxman,D.J. (2001). Serine phosphorylation of GH-activated signal transducer and activator of transcription 5a (STAT5a) and STAT5b: impact on STAT5 transcriptional activity. Mol Endocrinol *15*, 2157-2171.

Pena,R.N., Folch,J.M., Sánchez,A., and Whitelaw,C.B.A. (1998). Chromatin Structure of Goat and Sheep β-lactoglobulin Gene Differ. Biochemical and Biophysical Research Communications *252*, 649-653.

Pennings,S. (1997). Nucleoprotein Gel Electrophoresis for the Analysis of Nucleosomes and Their Positioning and Mobility on DNA. Methods: A Companion to Methods in Enzymology *12*, 20-27.

Pennings,S., Meersseman,G., and Bradbury,E.M. (1991). Mobility of Positioned Nucleosomes on 5S rDNA. J. Mol Biol *220*, 101-110.

Pennings,S., Meersseman,G., and Bradbury,E.M. (1992). Effect of glycerol on the separation of nucleosomes and bent DNA in low ionic strength polyacrylamide-gel electrophoresis. Nucleic Acids Research *20*, 6667-6672.

Pennock,G.D., Shoemaker,C., and Miller,L.K. (1984). Strong and regulated Expression of *Escherichia coli* β-galactosidase in Insect Cells with a Baculovirus Vector. Mol Cell Biol *4*, 399-406.

Pereira,S.L., Grayling,R.A., Lurz,R., and Reeve,J.N. (1997). Archaeal Nucleosomes. Proc Natl Acad Sci USA *94*, 12633-12637.

Perlmann,T. and Wrange,O. (1988). Specific Glucocorticoid Receptor-Binding to DNA reconstituted in a Nucleosome. EMBO J *7*, 3073-3079.

Pfitzner,E., Jahne,R., Wissler,M., Stoecklin,E., and Groner,B. (1998). P300/CREB-Binding protein enhances the prolactin-mediated transcriptional induction through direct interaction with the transactivation domain of Stat5, but does not participate in the Stat5-mediated suppression of the glucocorticoid response. Molecular Endocrinology *12*, 1582-1593.

Philp,J.A., Burdon,T.G., and Watson,C.J. (1996). Differential activation of Stats 3 and 5 during mammary gland development. Febs Lett *396*, 77-80.

Pierre,S., Jolivet,G., Devinoy,E., and Houdebine L-M. (1994). A Combination of Distal and Proximal regions Is Required for Efficient Prolactin Regulation of Transfected Rabbit αS1-

Casein Chloramphenicol Acetlytransferase Constructs. Molecular Endocrinology *8*, 1720-1730.

Piña,B., Brüggermeier,U., and Beato,M. (1990). Nucleosome Positioning Modulates Accessibility of Regulatory Proteins to the Mouse Mammary Tumor Virus Promoter. Cell *60*, 719-731.

Porath,J. (1992). Immobilized Metal Ion Affinity Chromatography. Protein Expression And Purification *3*, 263-281.

Porath,J., Carlsson,J., Olsson,I., and Belfrage,G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. Nature *258*, 598-599.

Prunell,A. (1983). Periodicity of Exonuclease III Digestion of Chromatin and the Pitch of Deoxyribonucleic Acid on the Nucleosome. Biochemistry *22*, 4887-4894.

Rascle,A., Johnston,J.A., and Amati,B. (2003). Deacetylase Activity Is required for Recruitment of the Basal Transcription Machinery and Transactivation by STAT5. Mol Cell Biol *23*, 4162-4173.

Rascle,A. and Lees,E. (2003). Chromatin acetylation and remodeling at the Cis promoter during Stat5-induced transcription. Nucleic Acids Research *31*, 6882-6890.

Rhodes,D. (1985). Structural analysis of a triple complex between the histone octamer, a *Xenopus* gene for 5 S RNA and transcription factor IIIA. EMBO J *4*, 3473-3482.

Richard-Foy,H. and Hager,G.L. (1987). Sequence-specific positioning of nucleosomes over the steroid-inducible MMTV promoter. EMBO J *6*, 2321-2328.

Richardson,C.C. and Kornberg,A. (1964). A Deoxyribonucleic Acid Phosphatase-Exonuclease from *Escherichia coli* I. J. Biol. Chem *239*, 242-250.

Richardson,C.C., Lehman,I.R., and Kornberg,A. (1964). A Deoxyribonucleic Acid Phosphatase-Exonuclease from *Escherichia coli* II. J. Biol. Chem *239*, 251.

Richmond,T.J. and Davey,C.A. (2003). The Structure of DNA in the Nucleosome Core. Nature *423*, 145-150.

Ridderstrale,M. and Groop,L. (2001). Differential phosphorylation of Janus kinase 2, Stat5A and Stat5B in response to growth hormone in primary rat adipocytes. Mol Cell Endocrinol *183*, 49-54.

Riley,D. and Weintraub,H. (1978). Nucleosomal DNA is Digested to Repeats of 10 Bases by ExonucleaseIII. Cell *13*, 281-293.

Robyr,D., Gegonne,A., Wolffe,A.P., and Wahli,W. (2000). Determinants of vitellogenin B1 promoter architecture - HNF3 and estrogen responsive transcription within chromatin. J. Biol. Chem *275*, 28291-28300.

Roskelley,C.D., Desprez,P.Y., and Bissell,M.J. (1994). Extracellular matrix-dependent tissue-specific gene expression in mammary epithelial cells requires both physical and biochemical signal transduction. Proc Natl Acad Sci USA *91*, 12378-12382.

Ruh,M.F., Chrivia,J.C., Cox,L.K., and Ruh,T.S. (2004). The interaction of the estrogen receptor with mononucleosomes. Molecular And Cellular Endocrinology *214*, 71-79.

Rusterholz,C., Henrioud,P.C., and Nabholz,M. (1999). Interleukin-2 (IL-2) regulates the accessibility of the IL-2- responsive enhancer in the IL-2 receptor alpha gene to transcription factors. Mol Cell Biol *19*, 2681-2689.

Sadowski,H.B., Shuai,Ke., Darnell,J.E.j., and Gilman,M.Z. (1993). A Common Nuclear Signal transduction Pathway activated by Growth Factor and Cytokine receptors. Science *261*, 1739-1744.

Schild,C., Claret,F.-X., Wahli,W., and Wolffe,A.P. (1993). A Nucleosome-Dependent Static Loop Potentiates Estrogen-Regulated Transcription from the *Xenopus* Vitellogenin B1 Promoter *in vitro*. EMBO J *12*, 423-433.

Schindler,U., Wu,P., Rothe,M., Brasseur,M., and McKnight,S.L. (1995). Components of a Stat Recognition code: Evidence for Two Layers of Molecular Selectivity. Immunity *2*, 689-697.

Schmitt-Ney,M., Doppler ,W., Ball,R.K., and Groner,B. (1991). β-Casein Gene Promoter Activity Is regulated by the Hormone-Mediated Relief of Transcriptional Repression and a Mammary-Gland-Specific Nuclear Factor. Mol Cell Biol *11*, 3745-3755.

Schmitt-Ney,M., Happ,B., Ball,R.K., and Groner,B. (1992). Developmental and environmental regulation of a mammary gland-specific nuclear factor essential for transcription of the gene encoding β–casein. Proc Natl Acad Sci USA *89*, 3130-3134.

Schnieke,A.E., Kind,A.J., Ritchie,W.A., Mycock,K., Scott,A.R., Ritchie,M., Wilmut,I., Colman,A., and Campbell,K.H.S. (1997). Human factor IX Transgenic Sheep produced by Transfer of Nuclei from Transfected Fetal Fibroblasts. Science *278*, 2130-2133.

Schroeder,M., Kroeger,K.M., Volk,H.-D., Eidne,K.A., and Grütz,G. (2004). Preassociation of nonactivated STAT3 molecules demonstrated in living cells using bioluminescence resonance energy transfer: a new model of STAT activation? Journal of Leukocyte Biology *75*, 729-797.

Schug,J. and Overton,G.C. (1997). TESS: Transcription Element Search Software on the WWW'. Computational Biology and Informatics Library. School of medicine. University of Pennsylvania *Technical report CBIL-TR-1997-1001-v0.0*.

Seidel,H.M., Milocco,L.M., Lamb,P., Darnell,J.E.J., Stein,R., and Rosen,J. (1995). Spacing of palindromic half sites as a determinant of selective STAT (signal transducer and activation of transcription) DNA binding and transcriptional activity. Proc Natl Acad Sci USA *92*, 3041-3045.

Shioda,M., Sugimori,T., and Takayanagi,S. (1989). Nucleosomelike Structures Associated with Chromosomes of the Archaebacterium *Halobacterium salinarium*. Journal of Bacteriology *171*, 4514-4517.

Shuai,K., Horvath,C.M., Huang,L.H., Qureshi,S.A., Cowburn,D., and Darnell,J.E.J. (1994). Interferon activation of the transcription fcator Stat91 involves dimerisation through SH2-phosphotyrosyl peptide interactions. Cell *76*, 821-828.

Shuai,Ke., Schindler,C., Prezioso,V.R., and Darnell,J.E.J. (1992). Activation of TRanscription by IFNγ: Tyrosine Phosphorylation of a 91-kD Binding protein. Science *258*, 1808-1812.

Simpson,R.T. (1978). Structure of the Chromatosome, a Chromatin Particle Containing 160 Base Pairs of DNA and All the Histones. Biochemistry *17*, 5524-5531.

Simpson,R.T. (1991). Nucleosome Positioning: Occurrence, Mechanisms and Functional Consequences. Progress in Nucleic Acid Research *40*, 143-184.

Simpson,R.T. and Stafford,D.W. (1983). Structural features of a phased nucleosome core particle. Proc Natl Acad Sci USA *80*, 51-55.

Smith,G.E., Summers,M.D., and Fraser,M.J. (1983). Production of Human Beta Interferon in Insect Cells Infected with a Baculovirus Expression Vector. Mol Cell Biol *3*, 2156-2165.

Socolovsky,M., Nam,H., Fleming,M.D., Haase,V.H., Brugnara,C., and Lodish,H.F. (2001). Ineffective erythropoiesis in Stat5a(-/-)5b(-/-) mice due to decreased survival of early erythroblasts. Blood *98*, 3261-3273.

Soldaini,E., John,S., Moro,S., Bollenbacher,J., Schindler,U., and Leonard,W.J. (2000). DNA binding site selection of dimeric and tetrameric Stat5 proteins reveals a large repertoire of divergent tetrameric Stat5a binding sites. Mol Cell Biol *20*, 389-401.

Soler-Lopez,M., Petosa,C., Fukazawa,M., Ravelli,R., Williams,J.G., and Muller,C.W. (2004). Structure of and Activated *Dictostelium* STAT in its DNA-Unbound Form. Molecular Cell *13*, 791-804.

Soulier,S., Iepourry,L., Stinnakre,M.-G., Langley,B., L'Huillier,P.J., Paly,J., Djiane,J., Mercier,J.-C., and Vilotte,J.L. (1999). Introduction of a proximal Stat5 site in the murine α-lactalbumin promoter induces prolactin dependency *in vitro* and improves expression frequency *in vivo*. Transgenic Research *8*, 23-31.

Spangenberg,C., Elsfeld,K., Stünkel.W., Luger,K., Flaus,A., Richmond,T.J., Truss,M., and Beato,M. (1998). The Mouse mammary Tumor Virus Promoter Positioned on a Tetramer of Histones H3 and H4 Binds Nuclear Factor 1 and OTF1. J. Mol Biol *278*, 725-739.

Strehlow,I. and Schindler,C. (1998). Amino-terminal signal transducer and activator of transcription (STAT) domains regulate nuclear translocation and STAT deactivation. J. Biol. Chem *273*, 28049-28056.

Streulli,C.H. and Edwards,G.M. (1998). Control of normal epithelial phenotype by integrins. Journal of Mammary Gland Biology and Neoplasia *3*, 151-163.

Struhl,K. (1998). Histone acetylation and transcriptional regulatory mechanisms. Genes and Development *12*, 599-606.

Struhl,K. (1999). Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. Cell *98*, 1-4.

Stünkel,W., Kober,I., and Seifart,K.H. (1997). A Nucleosome Positioned in the Distal Promoter Region Activates Transcription. Mol Cell Biol *17*, 4397-4405.

Sun,F.-L. and Elgin,S.C. (1999). Putting Boundaries on Silence. Cell *99*, 459-462.

Tam,S.P., Lau,J., Djiane,J., and Waters,M.J. (2001). Tissue-Specific Induction of SOCS Gene Expression by PRL. Endocrinology *142*, 5015-5026.

Taniuchi,H., Anfinsen,C.B., and Sodja,A. (1967). The Amino Acid Sequence of an Extracellular Nuclease of *Staphylococcus aureus*. J. Biol. Chem *242*, 4752-4758.

Tatchell,K. and van Holde,K.E. (1977). Reconstitution of Chromatin Core Particles. Biochemistry *16*, 5295-5303.

Tatchell,K. and van Holde,K.E. (1978). Compact oligomers and Nucleosome Phasing. Proc Natl Acad Sci USA *75*, 3583-3587.

Teglund,S., McKay,C., Scheultz,E., van Deursen,J.M., Stravopodis,D., Wang,D., Brown,M., Bodner,S., Grosveld,G., and Ihle,J.N. (1998). Stat5a and Stat5b Proteins Have Essential and Nonessential, or Redundant, Roles in Cytokine Responses. Cell *93*, 841-850.

Thoma,F. (1992). Nucleosome Positioning. Biochimica et Biophysica Acta *1130*, 1-19.

Thomas, J. O. and Kornberg, R. D. An Octamer of histones in hromatin and free in solution. Proceedings of the national Academy of Sciences USA 72[7], 2626-2630. 1975.

Udy,G.B., Towers,R.P., Snell,R.G., Wilkins,R., Park,S.-H., Ram,P.A., Waxman,D.J., and Davey,H. (1997). Requirement of STAT5b for Sexual Dimorphism of Body Growth Rates and Liver Gene Expression. Proc Natl Acad Sci USA *94*, 7239-7244.

Verdier,F., Rabionet,R., Gouilleux,F., Beisenherz-Huss,C., Varlett,P., Muller,O., Mayeux,P., Lacombe,C., Gisselbrecht,S., and Chretien,S. (1998). A sequence of the CIS gene promoter Interacts Preferentially with Two Associated STAT5A Dimers: A distinct Biochemical Difference Between STAT5A and STAT5B. Mol Cell Biol *18*, 5852-5860.

Vermaak,D., Ahmad,K., and Henikoff,S. (2003). Maintainance of chromatin states: an open-and-shut case. Current Opinion in Cell Biology *15*, 266-274.

Vilotte,J.L. and Soulier,S. (1992). Isolation and Characterization of the Mouse Alpha-Lactalbumin-Encoding Gene: Interspecies Comparison, Tissue- and Stage-Specific Expression. Gene *119*, 287-292.

Vincze,T., Posfai,I., and Roberts,R.I. (2003). NEBcutter: A Programme to cleave DNA with restriction Enzymes. Nucleic Acids Research *31*, 3688-3691.

Vinkemeier,U., Cohen,S.L., Moarefi,I., Chait,B.T., Kuriyan,J., and Darnell,J.E.J. (1996). DNA Binding of *in vitro* activated Stat1α, Stat1β and truncated Stat1: interaction between NH2-terminal domains stabilizes binding of two dimers to tandem DNA sites. EMBO J *15*, 5616-5626.

Vinkemeier,U., Moarefi,I., Darnell,J.E.j., and Kuriyan,J. (1998). Structure of the Amino-Terminal Protein Interacton Domain of STAT-4. Science *279*, 1048-1052.

Vitolo,J., Yang,Z., Basavappa,R., and Hayes,J.J. (2004). Structural Features of Transcription Factor IIIA Bound to a Nucleosome in Solution. Mol Cell Biol *24*, 697-707.

Volpe,T., Kidner,C., Hall,I.M., Teng,G., Grewal,S.I.S., and Martienssen,R.A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Wakao,H., Gouilleux,F., and Groner,B. (1994). Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor gene family and confers the prolactin response. EMBO J *13*, 2182-2192.

Wakao,H., Schmitt-Ney,M., and Groner,B. (1992). Mammary Gland-specific Nuclear factor Is Present in Lactating Rodent and Bovine Mammary Tissue and Composed of a Single Polypeptide of 89kDa. J. Biol. Chem *23*, 16365-16370.

Wang,D., Stravopodis,D., Teglund,S., Kitazawa,J., and Ihle,J.N. (1996). Naturally Occurring Dominant Negative Variants of Stat5. Mol Cell Biol *16*, 6141-6148.

Wang,M., Liu,Y.E., Goldberg,I.D., and Shi,Y.E. (2003). Induction of Mammary Gland Differentiation in Transgenic Mice by the Fatty Acid-Binding Protein MRG. J. Biol. Chem *278*, 47319-47325.

Watson,C.J., Gordon,K.E., Robertson,M., and Clark,A.J. (1991). Interaction of DNA-Binding proteins with a milk protein gene promoter in vitro: identification of a mammary gland-specific factor. Nucleic Acids Research *19*, 6603-6610.

Webster,J., Wallace,R., Clark,A.J., and Whitelaw,C.B.A. (1995). Tissue-Specific, Temporally Regulated Expression Mediated by the Proximal Ovine β-lactoglobulin Promoter in Transgenic Mice. Cellular and Molecular Biology research *41*, 11-15.

Weischet,W.O., Tatchell,K., van Holde,K.E., and Klump,H. (1978). Thermal denaturation of Nucleosomal Core particles. Nucleic Acids Research *5*, 139.

Whitelaw,C.B. and Webster,J. (1998). Temporal profiles of appearance of DNase 1 hypersensitive sites associated with the ovine beta-lactoglobulin gene differ in sheep and transgenic mice. Mol Gen Genet *257*, 649-654.

Whitelaw,C.B.A., Grolli,S., Accornero,P., Donofrio,G., Farini,E., and Webster,J. (2000). Matrix attachment region regulates basal β-Lactoglobulin transgene expression. Gene *244*, 73-80.

Whitelaw,C.B.A., Harris,S., McLenaghan,M., Simons,J.P., and Clark,A.J. (1992). Position-independent expression of the ovine β-lactoglobulin gene in transgenic mice. Biochem J *286*, 31-39.

Wilhelm,F.X., Wilhelm,M.L., and Daune,M.P. (1978). Reconstitution of chromatin: assembly of the nucleosome. Nucleic Acids Research *5*, 505-521.

Woelfle,J., Chia,D.J., and Rotwein,P. (2003). Mechanisms of Growth Hormone (GH) Action. J. Biol. Chem *278*, 51261-51266.

Woese,C.R. and Fox,G.E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc Natl Acad Sci USA *74*, 5088-5090.

Wolffe,A.P. (2001). Transcriptional regulation in the context of chromatin structure. Essays in Biochemistry *37*, 57.

Workman,J.L. and Kingston,R.E. (1998). Alteration of Nucleosome Structure as a Mechanism of Transcriptional Regulation. Annu Rev Biochem *67*, 545-579.

Wu,C. (1980). The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature *286*, 854-60.

Wyszomierski,S.L. and Rosen,J.M. (2001). Cooperative effects of STAT5 (signal transducer and activator of transcription 5) and C/EBP beta (CCAAT/enhancer-binding protein-beta) on beta-casein gene transcription are mediated by the glucocorticoid receptor. Molecular Endocrinology *15*, 228-240.

Xu,M., Nie,L., Kim,S.-H., and Sun,X.-H. (2003). STAT5-induced Id-1 transcription involves recruitment of HDAC1 and deacetylation of C/EBPβ. EMBO J *22*, 893-904.

Yamamoto,K., Shibata,F., Miyasaka,N., and Miura,O. (2002). The Human perforin gene is a direct target of STAT4 activated by Il-12 in NK cells. Biochem Biophys Res Commun *297*, 1245-1252.

Yamashita,H., Nevalainen,M.T., Xu,J., LeBaron,M.J., Wagner,K.U., Erwin,R.A., Harmon,J.M., Hennighausen,L., Kirken,R.A., and Rui,H. (2001). Role of serine phosphorylation of Stat5a in prolactin-stimulated beta-casein gene expression. Mol Cell Endocrinol *183*, 151-163.

Yan,R., Small,S., Desplan,C., Dearolf,C.R., and Darnell,J.E.J. (1996). Identification of a Stat gene that functions in Drosophila development. Cell *84*, 421-430.

Yenidunya,A., Davey,C., Clark,D., Felsenfeld,G., and Allan,J. (1994). Nucleosome positioning on chicken and human globin gene promoters in vitro. Novel mapping techniques. J. Mol Biol *237*, 401-414.

Zakharova,N., Lymar,E.S., Yang,E., Malik,S., Zhang,J.J., Roeder,R.G., and Darnell,J.E.J. (2003). Distinct Transcriptional Activation Functions of STAT1α and STAT1β on DNA and Chromatin templates. J. Biol. Chem *278*, 43067-43073.

Zhang,S., Fukuda,S., Lee,S., Hangoe,G., Cooper,S., Spolski,R., Leonard,W.J., and Broxmeyer,H.E. (2000). Essential Role of Signal Transducer and Activator of Transcription (Stat)5a but Not Stat5b for Flt-3 dependent signalling. The Journal of Experimental Medicine *192*, 719-728.

Zhang,X. and Darnell,J.E.J. (2001). Functional Importance of Stat3 Tetramerisation in Activation of the α2-Macroglobulin Gene. J. Biol. Chem *276*, 33576-33581.

Zhu,Z. and Thiele,D. (1996). A Specialized Nucleosome Modulates Transcription Factor Access to a C. glabrata Metal Responsive Promoter. Cell *87*, 459-470.

Zoubiane,G.S., Valentijn,A., Lowe,E.T., Akhtar,N., Bagley,S., Gilmore,A.P., and Streulli,C.H. (2003). A Role for the cytoskeleton in prolactin-dependent mammary epithelial cell differentiation. Journal of Cell Science *117*, 271-280.

Appendices

Alignment of Stat5a and Stat5b using CLUSTAL W (1.74). The conserved tyrosine involved in activation of the protein is highlighted in red.

```
Stat5a          MAGWIQAQQLQGDALRQMQVLYGQHFPIEVRHYLAQWIESQPWDAIDLDNPQDRAQATQL
Stat5b          MAVWIQAQQLQGEALHQMQALYGQHFPIEVRHYLSQWIESQAWDSVDLDNPQENIKATQL
                ** *********:**:***.**************:******.**::******:. :****

Stat5a          LEGLVQELQKKAEHQVGEDGFLLKIKLGHYATQLQKTYDRCPLELVRCIRHILYNEQRLV
Stat5b          LEGLVQELQKKAEHQVGEDGFLLKIKLGHYATQLQNTYDRCPMELVRCIRHILYNEQRLV
                ************************************:******:****************

Stat5a          REANNCSSPAGILVDAMSQKHLQINQTFEELRLVTQDTENELKKLQQTQEYFIIQYQESL
Stat5b          REANNGSSPAGSLADAMSQKHLQINQTFEELRLVTQDTENELKKLQQTQEYFIIQYQESL
                ***** ***** *.**************************************************

Stat5a          RIQAQFAQLAQLSPQERLSRETALQQKQVSLEAWLQREAQTLQQYRVELAEKHQKTLQLL
Stat5b          RIQAQFGPLAQLSPQERLSRETALQQKQVSLEAWLQREAQTLQQYRVELAEKHQKTLQLL
                ******. ******************************************************

Stat5a          RKQQTIILDDELIQWKRRQQLAGNGGPPEGSLDVLQSWCEKLAEIIWQNRQQIRRAEHLC
Stat5b          RKQQTIILDDELIQWKRRQQLAGNGGPPEGSLDVLQSWCEKLAEIIWQNRQQIRRAEHLC
                ************************************************************

Stat5a          QQLPIPGPVEEMLAEVNATITDIISALVTSTFIIEKQPPQVLKTQTKFAATVRLLVGGKL
Stat5b          QQLPIPGPVEEMLAEVNATITDIISALVTSTFIIEKQPPQVLKTQTKFAATVRLLVGGKL
                ************************************************************

Stat5a          NVHMNPPQVKATIISEQQAKSLLKNENTRNECSGEILNNCCVMEYHQATGTLSAHFRNMS
Stat5b          NVHMNPPQVKATIISEQQAKSLLKNENTRNDYSGEILNNCCVMEYHQATGTLSAHFRNMS
                ********************************: ****************************

Stat5a          LKRIKRADRRGAESVTEEKFTVLFESQFSVGSNELVFQVKTLSLPVVVIVHGSQDHNATA
Stat5b          LKRIKRSDRRGAESVTEEKFTILFESQFSVGGNELVFQVKTLSLPVVVIVHGSQDNNATA
                ******:***************:*********.************************:****

Stat5a          TVLWDNAFAEPGRVPFAVPDKVLWPQLCEALNMKFKAEVQSNRGLTKENLVFLAQKLFNN
Stat5b          TVLWDNAFAEPGRVPFAVPDKVLWPQLCEALNMKFKAEVQSNRGLTKENLVFLAQKLFNN
                ************************************************************

Stat5a          SSSHLEDYSGLSVSWSQFNRENLPGWNYTFWQWFDGVMEVLKKHHKPHWNDGAILGFVNK
Stat5b          SSSHLEDYSGLSVSWSQFNRENLPGRNYTFWQWFDGVMEVLKKHLKPHWNDGAILGFVNK
                ************************* ***************** ***************

Stat5a          QQAHDLLINKPDGTFLLRFSDSEIGGITIAWKFDSPERNLWNLKPFTTRDFSIRSLADRL
Stat5b          QQAHDLLINKPDGTFLLRFSDSEIGGITIAWKFDSQERMFWNLMPFTTRDFSIRSLADRL
                ********************************** ** :*** ****************

Stat5a          GDLSYLIYVFPDRPKDEVFSKYYTPV-----LAKAVDGYVKPQIKQVVPEFVNASADAGG
Stat5b          GDLNYLIYVFPDRPKDEVYSKYYTPVPCESATAKAVDGYVKPQIKQVVPEFVNASADAGG
                ***.**************:*******      ****************************

Stat5a          SSATYMDQAPSPAVCPQAPYNMYPQNPDHVLDQDGEFDLDETMDVARHVEELLRRPMDSL
Stat5b          GSATYMDQAPSPAVCPQAHYNMYPQNPDSVLDTDGDFDLEDTMDVARRVEELLGRPMD--
                .***************** ********* *** **:***::******:***** ****

Stat5a          DSRLSPPAGLFTSARGSLS
Stat5b          -SQWIPHAQS---------
                 *:  * *
```

Alignment of the Caprine and Ovine proximal promoter sequences using CLUSTAL W (1.74). The insertion is highlighted in yellow. The dyad axis mapped *in vitro* by (Boa, 1999) for each sequence is highlighted in pink.

```
Caprine         --------------------TTGGAGGAGCTGGTGCCCAAGGCAGAGGCCACCCTCCAGGA
Ovine           --------------------TTGGAGGAGCTGGTGCCCAAGGCAGAGGCCACCCTCCAGGA
                                    ****************************************

Caprine         CACACCTGTCCCCAGTGCTGGCTCTGACCTGCCCTTGTCTAAGAGGCTGACCCCGGAAGT
Ovine           CACACCTGTCCCCAGTGCTGGCTCTCAGCTGTCCTTGTCTAAGAGGCTGACCCCGGAAGT
                ************************* * *** ****************************

Caprine         GTTCCTGGCACTGGCAGCCAGCCTG-ACCCAGAGTCCAGACACCCACCTGTGCCCCCACT
Ovine           GTTCCTGGCACTGGCAGCCAGCCTGGACCCAGAGTCCAGACACCCACCTGTGCCCCCGCT
                ************************* ******************************** **

Caprine         TCTGGGGTCTACCAGGAACCGTCTAGGCCCAGAGGGGGACTTCCTGCTTGGCCCCGGATG
Ovine           TCTGGGGTCTACCAGGAACCGTCTAGGCCCAGAGGGGGACTTCCTGCTTGGCCTTGGATG
                ********************************************************  *****

Caprine         GAAGAAGGCCTCCTATTGTCCTCGTAGAGGAAGCCACCCCGGGGCCCGGGGATGAGCCAA
Ovine           GAAGAAGGCCTCCTATTGTCCTCGTAGAGGAAGCCACCCCGGGGCCTGAGGATGAGCCAA
                ********************************************** * ***********

Caprine         GTAGGATTCCGGGAACCTCGTGGCTGGGGGCCCGGCCCGGGCTGGCTGGCTGGCACGCCT
Ovine           GTGGGATTCCGGGAACCGCGTGGCTGGGGTACCAGCCCGGGCTGGCTGGCCTGCATGCCT
                ** ************** ***********  ** ****************  *** ****

Caprine         CCTGTATAAGGCCCCGAGCCCGCTGTCTCAGCCCTCCACTCCCTGCAGAGCTCAGAAGCA
Ovine           CCTGTATAAGGCCCCAAGCCTGCTGTCTCAGCCCTCCACTCCCTGCAGAGCTCAGAAGCA
                *************** **** ****************************************

Caprine         CGACCCCAGCTGCAGCCATGAAGTGCCTCCTGCTTGCCCTGGGCCT
Ovine           CGACCCCAGCTGCAGCCATGAAGTGCCTCCTGCTTGCCCTGGGCCT
                ***********************************************
```